

***Spatialization* von nutzergenerierten Inhalten für die explorative
Analyse des globalen Städtetetzes**

D i s s e r t a t i o n

z u r

Erlangung der naturwissenschaftlichen Doktorwürde
(*Dr. sc. nat.*)

v o r g e l e g t d e r

Mathematisch-naturwissenschaftlichen Fakultät
d e r
Universität Zürich

v o n

Marco Michele Salvini

v o n

Cama (GR)

P r o m o t i o n s k o m i t e e

Prof. Dr. Sara Irina Fabrikant
(Vorsitz & Leitung)

Prof. Dr. Céline Rozenblat

Dr. Tumasch Reichenbacher

Zürich, 2012

Vorwort

Die vorliegende Arbeit ist zwischen 2009 und 2012 in der Abteilung Geographische Informationsanalyse und –visualisierung (GIVA) am Geographischen Institut der Universität Zürich unter der Leitung von Prof. Dr. Sara I. Fabrikant entstanden.

Das Geographische Institut war nach dem Studium (Bachelor & Master) in Wirtschaftsgeographie und Raumplanung und der Anstellung als Studienberater kein Neuland. Dennoch öffnete sich mir durch die neue Herausforderung eine noch fast unbekannte wissenschaftliche Dimension.

Meine neuen ArbeitskollegInnen beschäftigten sich mit geographischer Relevanz, Entscheidungen mit Karten unter Zeitdruck, empirischer Evaluation kognitiver Aspekte der (Geo)Visualisierung und *geofovation*. Dieser zu Beginn „fremde Nährboden“ hat sich als besonders fruchtbar für den interdisziplinären Austausch gezeigt und als eine riesige Chance erwiesen, um geoanalytische und kartographische Forschungsaspekte zu vertiefen. Die vorliegende Dissertation positioniert sich inhaltlich an der Schnittstelle zwischen meinem wirtschafts- und stadtgeographischen Hintergrund und dem neuen geoinformationswissenschaftlichen Umfeld.

Ich möchte an dieser Stelle allen danken, mit denen ich diese Zeit teilen durfte und die mich unterstützt haben.

Der grösste Dank geht an Prof. Dr. Sara I. Fabrikant für ihre engagierte Betreuung und, dass sie mir ermöglichte, mich in der Forschung und in der Lehre weiterzuentwickeln. Die grosse wissenschaftliche Freiheit und der wertvolle fachliche sowie persönliche

Austausch haben massgeblich zum Gelingen der vorliegenden Arbeit beigetragen. Danken möchte ich auch den Mitgliedern des Promotionskomitees, Prof. Dr. Céline Rozenblat und Dr. Tumasch Reichenbacher, für die kompetente Beratung und die wertvollen Kommentare zur Arbeit.

Die Leitung der Übungen des Kurses Kartographie und Visualisierung hat die Zeit der Doktorarbeit massgeblich geprägt: Der Austausch mit den Studierenden, den TutorInnen und den Assistierenden waren eine willkommene Abwechslung im Forschungsalltag. Ich danke dabei Simon Etter, Kaspar Fischer, Michael Pichlmeier und Martin Zahner für die hervorragende Zusammenarbeit, die über das Berufliche hinaus geht.

Ein grosser Dank gilt meiner Mutter, Dorothea Salvini-Kim, sowie André Bruggmann für das präzise und geduldige Korrekturlesen der Arbeit. Zudem möchte ich mich bei den KollegInnen der Abteilung, bei den MitarbeiterInnen der zentralen Dienste des Geographischen Instituts und bei den Mitarbeiterinnen des Studiendekanats MNF, die mich in verschiedenster Art und Weise unterstützt haben, bedanken.

Bei Nicola Kugelmeier und Amalia Schneider, sowie bei meinen langjährigen Mitbewohnern Fabio Ferretti und Lorenzo Oleggini möchte ich mich für ihre Begleitung im beruflichen und privaten Alltag bedanken. Ein besonderer Dank geht an Ursina Stoll für das Teilen der Abschlussphase dieser Arbeit. Meiner Familie möchte ich herzlich für die vorbehaltlose Unterstützung während des ganzen Studiums danken.

Zürich, 31. Juli 2012

Marco Salvini

Inhalt

Inhaltsverzeichnis

Vorwort	I
Inhalt	III
Inhaltsverzeichnis	III
Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
Formelverzeichnis	IX
Zusammenfassung	X
Summary	XII
1 Einführung	1
2 Stand der Forschung	9
2.1 Theoretische Ansätze	9
2.2 Etablierte Datengrundlagen und methodische Ansätze	18
2.3 Überblick der empirischen Ergebnisse	31
2.4 Fazit	40

3	Methodentheoretischer Rahmen	42
3.1	<i>Spatialization framework</i>	42
3.1.1	Semantische Generalisierung	46
3.1.2	Mathematische Transformation	52
3.1.3	Geometrische Generalisierung	55
3.1.4	Fazit: Grammatik der Netzwerkvisualisierung	57
3.2	Methodische Umsetzung	59
4	Datengrundlage	66
5	Methodik	72
5.1	Modellierung des allgemeinen Stdtenetzes	72
5.2	Analyse des allgemeinen Stdtenetzes	76
5.3	Differenzierung der Funktionen	80
5.4	Analyse des funktionalen Stdtenetzes	88
5.5	Evaluation	89
5.6	Fazit	95
6	Methodenevaluation	100
6.1	Datengrundlage	100
6.2	Verrumlichung mittels DrL/VxOrd-Algorithmus	110
6.3	Definition der Funktionen	117
7	Resultate - Fallstudien	126
7.1	Das globale Stdtenetz	126
7.2	Funktionen im Stdtenetz	145
7.3	Funktionale Stdtenetze	155
8	Diskussion	170
8.1	Datengrundlage	170
8.2	Methodischer Ansatz	176
8.2.1	Methodik	177
8.3	Globales Stdtenetz	183
9	Schlussfolgerungen	194
10	Bibliographie	200
11	Anhang	214
	Curriculum Vitae	226

Abbildungsverzeichnis

Abbildung 1:	Forschungsansatz der vorliegenden Arbeit	6
Abbildung 2:	Thematische Übersicht der Arbeit. Methodik nach Fabrikant und Salvini (2011)	8
Abbildung 3:	Urbanes Drei-Skalen-System nach Pumain (2006a) und Rozenblat (2010)	11
Abbildung 4:	Kartogramm mit der relativen <i>global connectivity</i> gewählter Städte (Taylor et al. 2002b)	21
Abbildung 5:	Urbane <i>arenas</i> (Derudder et al. 2003)	22
Abbildung 6:	Unterschiedliche Modellierungen des Städtetetzes (Rozenblat und Pumain 2007)	23
Abbildung 7:	Annahmen für die Modellierung des Städtetetzes (Rozenblat 2004)	25
Abbildung 8:	Globales Städtesystem im Jahr 2007 nach Alderson et al. (2010)	27
Abbildung 9:	Zentralität der europäischen Städte als a) kartographische (links) und b) schematische Darstellung (rechts) (Rozenblat 2004)	28
Abbildung 10:	Übersicht des methodentheoretischen Rahmens mit den verschiedenen Prozessschritten und den Resultaten, die sich daraus ergeben	46
Abbildung 11:	Titelseite der Dezember 2006 Ausgabe von <i>Time</i>	67
Abbildung 12:	Mögliche Relationen zwischen zwei Artikeln in Wikipedia. A) Mit gleicher Kategorie markiert. B) Einkommender Hyperlink aus gemeinsamem Artikel. C) Direkter Hyperlink von einem Artikel zum anderen. D) Ausgehender Hyperlink zu gemeinsamem Artikel	73
Abbildung 13:	Umwandlung des Artikelsystems in Wikipedia zu einem unimodalen Stadt-Stadt-Netzwerk	74
Abbildung 14:	Rollendefinition aufgrund von Z-Werten und P-Werten (Guimerà und Amaral 2005b)	79
Abbildung 15:	Typisierung der verbindenden Artikel aufgrund der Markierung mit Kategorien	81
Abbildung 16:	Einbezug der Artikel, Hyperlinks und Kategorien für die Bestimmung von funktionalen Beziehungen zwischen zwei Städten	88
Abbildung 17:	Arbeitsschritte für die Transformation der rohen Datengrundlage (Wikipedia) in ein Städtetetz	96
Abbildung 18:	Arbeitsschritte zur Visualisierung und Analyse des globalen Städtetetzes	97
Abbildung 19:	Arbeitsschritte zur Visualisierung und Analyse der Funktionen	99
Abbildung 20:	Städtetetz basierend auf der englischsprachigen Wikipedia. Dargestellt sind 127 von 516 Städten und der MST; periphere Städte sind nicht dargestellt	102
Abbildung 21:	Städtetetz basierend auf der englischen, deutschen, französischen und italienischen Wikipedia. Dargestellt sind 138 von 516 Städte und der MST; periphere Städte sind nicht dargestellt	103

Abbildung 22:	Scatterplot mit Regressionsgerade zwischen dem gewichteten <i>degree</i> im englischsprachigen und im Referenzstädtenetz	105
Abbildung 23:	Scatterplot mit Regressionsgerade zwischen dem Z-Wert im englischsprachigen und Referenzstädtenetz	106
Abbildung 24:	Scatterplot mit Regressionsgerade zwischen dem <i>degree</i> im MST des englischsprachigen und des Referenzstädtenetzes	107
Abbildung 25:	Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 500 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)	111
Abbildung 26:	Städte im geographischen Raum mit Hervorhebung der Referenzregionen	112
Abbildung 27:	Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 100 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)	113
Abbildung 28:	Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 15 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)	114
Abbildung 29:	Scatterplot mit Regressionsgeraden zwischen den intraregionalen Distanzen in drei Regionen im projizierten Raum (15 nächsten Nachbarn) und im Referenzraum	116
Abbildung 30:	Verräumlichung der Funktionen mit der Themenbildung anhand des Wikipedia-Kategoriensystems mittels einer SOM	118
Abbildung 31:	Projektion der mit <i>topic modelling</i> definierten Funktionen auf die nach dem Wikipedia-Kategoriensystem berechnete Geometrie	119
Abbildung 32:	Projektion der relationalen Regionen im geographischen Raum	127
Abbildung 33:	Regionen des globalen Städtensetzes im topologischen Raum	130
Abbildung 34:	Regionen des globalen Städtensetzes im topologischen Raum: Ausschnitt Europa und Zentralasien	131
Abbildung 35:	Regionen des globalen Städtensetzes im topologischen Raum: Ausschnitt südliche Welthalbkugel	133
Abbildung 36:	Regionen des globalen Städtensetzes im topologischen Raum: Ausschnitt Asien	134
Abbildung 37:	Zentralitätsanomalien zwischen <i>betweenness</i> Zentralität und gewichtetem <i>degree</i> der Städte nach Guimerà et al. (2005)	137
Abbildung 38:	Verteilung der Rollen im Städtensetz	138
Abbildung 39:	Globales Städtensetz zwischen den Hubs und deren Lage in der regionalen Ansicht	140
Abbildung 40:	Beziehungen der globalen Städte nach Regionen	143

Abbildung 41:	Funktionen im Städtenetz und deren Grösse	146
Abbildung 42:	Funktionen im Städtenetz und deren Grösse: Auszug WIRTSCHAFT/TECHNOLOGIE	148
Abbildung 43:	Verteilung der latenten Themen und der Regionen mit Akteuren im topologischen Raum	150
Abbildung 44:	Funktionen im Städtenetz und deren Grösse: Auszug POLITIK	152
Abbildung 45:	SOM mit zwei Ebenen: Funktionen als hierarchisch organisierte Regionen und Städte als Punktsymbole	156
Abbildung 46:	Komponentenräume der elf Funktionen und Subfunktionen	158
Abbildung 47:	Multifunktionales globales Städtenetz zwischen den Hubs	160
Abbildung 48:	Globales Städtenetz zwischen den Hubs aufgrund der Funktion POLITIK	162
Abbildung 49:	Globales Städtenetz zwischen den Hubs aufgrund der Subfunktion WIRTSCHAFT	164
Abbildung 50:	Globales Städtenetz zwischen den Hubs aufgrund der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR (Subfunktionen MILITÄR/KONFLIKTE und GESCHICHTE/-GESELLSCHAFT)	166
Abbildung 51:	Log-Log-Diagramm des gewichteten <i>degree</i> im MST die drei funktionale Netzwerke Funktionen und das allgemeinen Städtenetz	168

Tabellenverzeichnis

Tabelle 1:	Typologie der Beziehungen im Städtesystem nach Smith und Timberlake (1995a)	13
Tabelle 2:	Datensätze für die Modellierung multifunktionaler Netzwerke (Taylor 2005a)	30
Tabelle 3:	Ansätze zur Untersuchung des globalen Städtenetzes	32
Tabelle 4:	<i>Global cities</i> und <i>world cities</i> in Europa	34
Tabelle 5:	<i>Global cities</i> und <i>World cities</i> in Nordamerika	35
Tabelle 6:	<i>Global cities</i> und <i>world cities</i> in Asien	36
Tabelle 7:	<i>Global cities</i> und <i>world cities</i> in Ozeanien, Süd-, Zentralamerika und Afrika	37
Tabelle 8:	Der Kern des globalen Städtenetzes	38
Tabelle 9:	Taxonomie der Städte unter Berücksichtigung der wirtschaftlichen, politischen, gesellschaftlichen und kulturellen Funktionen (Anpassung von Taylor 2005a)	39
Tabelle 10:	Primitiven des Zielbereichs aus netzwerkanalytischer Perspektive und aus Sicht der <i>spatialization</i>	47
Tabelle 11:	Metaphorische Kartierung. Assoziierung des Zielbereichs (netzwerkanalytische und semantische Primitiven) mit dem Quellbereich definiert anhand der geographischen Primitiven in Navigationsperspektive nach Lynch (1962), den Image-Schemas nach	

	Johnson und Lakoff (Lakoff und Johnson 1980; Lakoff 1987; Johnson 1987) und den Primitiven des räumlichen Wissens nach Golledge (1995)	48
Tabelle 12:	Konzepte und Masse der Netzwerkanalyse unterteilt nach der Ebene der Analyse	54
Tabelle 13:	Grammatik der graphischen Darstellung	58
Tabelle 14:	Aus der englischen Wikipedia extrahierte und in der Arbeit verwendete Daten	71
Tabelle 15:	Hauptthemen in Wikipedia definiert anhand der Kategorie „ <i>Main topic classifications</i> “	82
Tabelle 16:	Einfache Kennzahlen der vier grössten Versionen von Wikipedia (englischsprachig, deutschsprachig, französischsprachig und italienischsprachig) nach Zachte (2011)	101
Tabelle 17:	Korrelationsanalyse zwischen dem Referenz- und dem englischsprachigen Städtenetz aufgrund ausgewählter Netzwerkmasse	104
Tabelle 18:	Vergleichsmatrix zwischen der regionalen Organisation im Referenz- und im englischsprachigen Städtenetz	108
Tabelle 19:	Evaluation der Robustheit des Datensatzes hinsichtlich A) der Volatilität der Einträge, B) der Modellierung der Stärken der Beziehungen zwischen den Städten und C) der Anzahl betrachteter Städte.	110
Tabelle 20:	Genauigkeitswerte für die drei betrachteten Projektionen	117
Tabelle 21:	Übereinstimmung der Funktionen nach Wikipedia (Referenzmodell) mit den Funktionen gemäss der <i>topic modelling</i> -Methode	120
Tabelle 22:	Vergleich der Relevanz- (tf-idf) und der absoluten Häufigkeitswerte relevanter Begriffe pro Funktion zwischen der Klassierung mit dem Wikipedia-Kategoriensystem und mit der <i>topic modelling</i> -Klassierung (TM)	122
Tabelle 23:	Zentralitäten (normalisierte Werte mit Rangwerten in Klammern) der zentralsten Städte	136
Tabelle 24:	Vergleich der absoluten und relativen Häufigkeiten von Akteuren pro Funktion und Subfunktion zwischen einem Datensatz mit zufällig ausgewählten Artikeln und dem Städtenetz	154
Tabelle 25:	Vergleich der Eigenschaften von Wikipedia mit den etablierten Datengrundlagen.	171
Tabelle 26:	Vergleich des Informationsgehalts verschiedener Darstellungen des globalen Städtenetzes	180
Tabelle 27:	Gegenüberstellung der Resultate der Arbeit und des Forschungsstandes. Resultate: Rangierung der 95 Hubs im allgemeinen und im wirtschaftlich/technologischen Netzwerk. Stand der Forschung: Kategorisierung und Rangierung der 95 Hubs nach den gleichen Quellen und Grundsätzen wie in Kapitel 2.3 (siehe Tabelle 3, Anhang 9 und Anhang 1)	186

Formelverzeichnis

Formel 1:	Transformation des bimodalen Netzwerks in ein Stadt-Stadt-Netzwerk mit Anwendung des <i>max-flow-min-cut</i> -Theorems	75
Formel 2:	Berechnung des normalisierten Kantengewichtes in Anlehnung an Hecht und Raubal (2008)	75
Formel 3:	Gewichteter <i>degree</i> für einen Knoten (n_i).	76
Formel 4:	Modularität für einzelne Subgruppen nach Newman und Girvan (2004) angepasst für gewichtete Netzwerke und normalisiert	78
Formel 5:	Z-Wert A) und P-Wert B) für die Bestimmung der Rollen (Guimerà und Amaral 2005b)	79
Formel 6:	Maximale relative Häufigkeit der Funktionen pro Neuron.	86
Formel 7:	Relevanz eines Begriffs gemessen mit der tf-idf-Methode (Manning et al. 2008)	87
Formel 8:	Einfache Berechnung des normalisierten Kantengewichts.	92
Formel 9:	Relative Übereinstimmung einer Funktion nach Wikipedia (FWI) und nach <i>topic modelling</i> (FTM)	94

Zusammenfassung

Im Zuge des Globalisierungsprozesses und der technologischen Entwicklung werden Informationen und Güter immer öfters und über grössere Distanzen ausgetauscht. Die Personenmobilität folgt demselben Trend. Die Flüsse des globalen Austausches sind ein prägendes Element unseres Zeitalters und bestimmen gleichwohl die Dynamik der Städte sowie ihre Integration zu einem globalen Städtenetz.

Auf theoretischer Ebene wurden mehrere Ansätze zu diesem Thema entwickelt. Die bekanntesten sind der *global cities*-Ansatz von Sassen (Sassen 1991) und die *world cities* Theorie von Friedmann (1986). Aufgrund der Schwierigkeit, globale und relationale Daten, anhand deren das globale Städtenetz erkundet werden kann, zu finden, ist der konzeptionelle Reichtum bei den empirischen Analysen nicht wiederzufinden. Die etablierten Ansätze beschränken sich mehrheitlich auf ökonomische Netzwerke. Die kulturellen, sozialen und politischen Funktionen werden nicht berücksichtigt.

Die vorliegende Arbeit greift diese Lücke auf und exploriert das Potential von nutzergenerierten Daten (im Konkreten die englischsprachige Version von Wikipedia) für die Modellierung und Analyse eines multifunktionalen und globalen Städtenetzes. Dafür wird ein interdisziplinärer methodischer Rahmen entwickelt, welcher auf der Triangulation von Netzwerkanalyse und -visualisierung, semantischer Analyse und Methoden der *spatialization* basiert. In einem ersten Teil werden ein theoretisch fundierter konzeptioneller methodischer Rahmen und zwei konkrete Arbeitsabläufe für die visuell-analytische Exploration und Analyse von Netzwerken im Allgemeinen entwickelt. Dar-

auffolgend werden diese für die Erkundung des globalen Städtensetzes konkretisiert. Zudem wird die Sensitivität und die Validität sowohl der nutzergenerierten Datengrundlage als auch der wichtigsten methodischen Schritte systematisch evaluiert.

Die Resultate sind sowohl methodischer als auch thematischer Natur. Erstens erweist sich die englische Version von Wikipedia als inhaltlich valide Datengrundlage für den betrachteten stadtgeographischen Forschungsbereich. Die Evaluation der potentiellen Risiken, die mit der wissenschaftlichen Nutzung einer *crowdsourced*, nutzergenerierten Datengrundlage verbunden sind, beeinträchtigen die Resultate nicht. Die Evaluation zeigt zudem, dass die Methoden zur Verräumlichung der Daten die latenten Muster und Strukturen der Daten genau und kohärent reproduzieren. Anhand der gewählten semantischen Ansätze lassen sich mehrere inhaltlich kohärente Funktionen identifizieren.

Drei Fallstudien legen die thematischen Ergebnisse dar. Die erste Fallstudie untersucht die Struktur des globalen Städtensetzes. Dabei heben sich London, New York und Paris wegen ihrer überdurchschnittlichen Zentralität hervor. Die zweite Fallstudie widmet sich der Definition der Funktionen. Für den niedrigsten Detaillierungsgrad ergeben sich acht Funktionen: PERSONEN, GEBAUTE/NATÜRLICHE UMWELT, POLITIK, WIRTSCHAFT/TECHNOLOGIE, SPORT, BILDUNG, GESELLSCHAFT/GESCHICHTE/MILITÄR und KUNST. In der letzten Fallstudie wird die Struktur der funktionalen Städtensetze untersucht. Daraus zeigt sich, dass einerseits die latente Grundstruktur aller untersuchten funktionalen Netzwerke ähnlich ist, andererseits sind funktionsspezifische strukturelle Nuancen erkennbar. So weisen etwa die EUROPÄISCHEN Städte eine ausgeprägte Zentralität in der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR auf, während sich die OST-/SÜDOSTASIATISCHEN Städte stärker in der Funktion WIRTSCHAFT/TECHNOLOGIE profilieren.

Zusammenfassend leistet diese Dissertation drei Beiträge. Erstens exploriert sie das Potential von nutzergenerierten Web 2.0 Inhalte für wissenschaftliche Zwecke; zweitens wird ein interdisziplinärer methodischer Rahmen für die für systematische Untersuchung von Netzwerkdaten im Allgemeinen und vom globalen Städtensetz im Spezifischen entwickelt; drittens werden dabei neue Erkenntnisse über die multifunktionale Einbettung der Städte im globalen Städtensetz generiert.

Summary

Castells (1996) identifies the space of flows as a new spatial form with a dominant role in the network society. On the urban level this space of flows is constituted by the exchanges of people, goods, and information between cities and it is an expression of the integration of the cities in an urban system. This functional network shapes and transforms the cities. The identification and the measurement of the functional network is one of the big challenges in this urban geographic research domain. The established approaches are mostly limited to the economic investigation of the global city networks. The cultural, social, and political functions in the network have never been adequately explored until now.

Motivated by the lack of data to study these processes this thesis investigates the potential of massively crowdsourced, user-generated databases (e.g., the English version of Wikipedia) for the systematic analysis of the global city network, and in particular, the space of flows of within these systems. The Wikipedia database is explored structurally and semantically, and analyzed with the integration of theoretically sound quantitative methods from social science (i.e., social network analysis, computational linguistics), coupled with computational approaches in GIScience and computer science (i.e., cartography, spatialization, graph drawing). In the first part of the thesis, a theoretically rooted conceptual framework, and two systematical approaches for the explorative network analysis are developed. Subsequently, the framework is applied to the exploration of the global city network extracted from the Wikipedia database. In addition, the sensi-

tivity and the validity of the user-generated database and of the most important methodological steps are systematically evaluated.

The results are both of methodological and thematic nature. First, the English version of Wikipedia appears to be a valid database for this considered urban geographic research approach. The systematic evaluation of the user-generated database suggests that the potential risks associated with the scientific use of Wikipedia does not affect or limit the quality of the results. Second, the spatialization methods reproduce coherent and meaningful patterns and latent structures in the data.

The framework is put to a rigorous test in three case studies to uncover the latent structure of the global city network. The first case study explores the structure of this network. London, New York, and Paris are identified as the core of the network. The second case study deals with the definition of the functions in the network. Different levels of details reveal different patterns. At the deepest level of detail there are eight functions: PEOPLE, BUILT/NATURAL ENVIRONMENT, POLITICS, ECONOMY/TECHNOLOGY, SPORTS, EDUCATION, SOCIETY/HISTORY/MILITARY, and ART. In the last case study, the functional city networks are further explored. On the one hand, the basic structure of the global city network is independent of the considered functions; on the other hand, functionally dependent patterns are recognizable: for example, the European cities have a higher centrality in the functions SOCIETY/HISTORY/MILITARY and ART, whereas the east-south-east Asian cities have higher centrality in the ECONOMY/TECHNOLOGY function.

In summary, this thesis has three main contributions. First, this thesis explores the potential of Web 2.0 user-generated content for scientific purposes. Second, an interdisciplinary methodological framework for the analysis of network data and specifically of the global city network is developed. Third, new insights about the multifunctional integration of the cities in the global city network are generated.

1 Einführung

Die gegenwärtige Gesellschaft ist von einer ökonomischen, kulturellen und gesellschaftlichen Globalisierung sowie von einer immensen technologischen Entwicklung geprägt. Castells (1996) bezeichnet sie als Informationsgesellschaft. Trotz der technologischen Möglichkeiten hat sich die Informationsgesellschaft nicht zu einer ubiquitären, raumlosen Gesellschaft entwickelt. Vielmehr sind Netzwerke, die sich zwischen bestimmten fokalen Punkten aufspannen, die entscheidende Organisationsform (Castells 2001). Städte sind die räumlichen Ankerpunkte dieser Netzwerke und somit fokale Orte unserer Gesellschaft. Die greifbarsten Aspekte dieser Netzwerke sind der Gütertransport, die Personenmobilität, der Informationsaustausch oder das globale Netz an Niederlassungen transnationaler Unternehmen. Die Städte werden dadurch in ein Städtesystem integriert, was die Entwicklung der Städte prägt. Die Städte, als Teil des Städtesystems, lassen sich damit „durch die Netzwerke erfassen, die [sie] durchziehen und bestimmen“ (Diener et al. 2006).

Die zentrale Rolle der Städte in der Gesellschaft ist persistent über die Zeit und bereits in anderen Zeitaltern präsent. Neu ist jedoch die Stärke, mit der sich die Städte in ein übergeordnetes Städtesystem integrieren. Neu ist auch die räumliche Skala des Systems. Während die Städte in der Vergangenheit in regionale oder nationale Städtesysteme eingebettet waren, sind Städtesysteme heute globaler Natur. In dieser Hinsicht spricht man von Weltstädten, von *global cities* oder *world cities* und bezieht sich auf Städte, deren Bedeutung weit über die nationalen Grenzen hinaus reicht. London und New York sind zwei prototypische Beispiele dafür.

Das Aufkommen der Netzwerkgesellschaft bringt die Stadt in eine dichotomische Stellung in Bezug auf die Konzepte von Raum und Nähe. Eine Stadt als geographische Entität befindet sich im geographischen Raum in einer bestimmten geographischen Distanz zu anderen Städten. So ist die zu überwindende geographische Distanz zwischen Paris und Berlin kleiner als zwischen Paris und Shanghai. Die Netzwerkstruktur bringt ein relationales Raumverständnis und ein weiteres Konzept von Distanz zum Vorschein: die topologische Distanz. Die topologische Distanz wird durch die Stärke der Integration der Städte ins Netzwerk und die Intensität der Beziehungen zwischen den Städten gegeben. In topologisch-relationaler Hinsicht können auch geographisch weit entfernte Städte nahe beieinander liegen. Die geographische Lage wird zu einem Attribut der Stadt analog zur Bevölkerungszahl oder zur wirtschaftlichen Wertschöpfung.

Die vorliegende Arbeit beschäftigt sich mit der Stadt als Knoten eines sozio-ökonomischen Systems (das Städtenetz oder Städtesystem), in der ihre Bedeutung in erster Linie durch relationale bzw. topologische und nur in zweiter Linie durch geographische Eigenschaften gegeben ist.

Problemstellung

Der theoretische Reichtum im Forschungsbereich der globalen Städte ist von einer empirischen Armut begleitet (Taylor 2004). Auf theoretischer Ebene kommt insbesondere der relationale bzw. netzwerkartige und der grenzüberschreitende bzw. globale Charakter der Urbanisation hervor. Die Identifikation einer Datengrundlage, welche diese Eigenschaften nachbilden könnte, und die Modellierung der Netzwerke stellen weiterhin eine besondere Herausforderung dar. In der Literatur wird oft und vehement auf diese Schwäche hingewiesen; so sprechen Beaverstock et al. (2000b) von einer „*Achillesferse*“ und Short et al. (1996) von „*The dirty little secret of world cities research*“. Verschiedene Forschungsprojekte sind im letzten Jahrzehnt insbesondere dieses Problem angegangen. Das globale Städtenetz wird aufgrund von Intrafirmenbeziehungen von *advanced producer services* (APS) (Taylor 1997; Taylor et al. 2002a, 2002b), aufgrund von Besitzverhältnissen in transnationalen Unternehmen (Rozenblat 2004; Wall und Van der Knaap 2011) oder aufgrund von Flugpassagiervolumen (Smith und Timberlake 1995a, 2001) modelliert. Diese Ansätze werden zwar einem Teil der gestellten Anforderungen gerecht, vermögen jedoch nicht alle theoretischen Aspekte empirisch umzusetzen. Die ersten zwei Ansätze beziehen sich ausschliesslich auf ökonomische Prozesse; gesellschaftliche, politische und kulturelle Aspekte bleiben unberücksichtigt. Die Flugpassagierdaten umfassen ein breiteres Spektrum an Interaktionen, die jedoch funktional undifferenziert sind.

Auf analytischer Ebene postulieren Smith und Timberlake (1995b) aufgrund des relationalen Charakters der Prozesse netzwerkanalytische Ansätze zur Exploration und Analyse des globalen Städtenetzes. Diesem Postulat wird zunehmend Rechnung getragen; Methoden der sozialen Netzwerkanalyse werden zunehmend eingesetzt. Das Potential insbesondere von (geo)visuell-analytischen Methoden bleibt jedoch weitgehend unausgeschöpft.

Vor diesem Hintergrund stellt sich die Problemstellung dieser Arbeit folgendermassen:

Das Generieren von neuem Wissen über das globale Städtenetz bedarf einer den theoretischen Feinheiten entsprechende Datengrundlage. Erstens muss diese eine grosse Anzahl über die ganze Welt verteilte Städte umfassen. Zweitens muss sie sich auf die Interaktionen zwischen den Städten und nicht auf die Eigenschaften der Städte beziehen. Drittens muss sie funktional umfassend und differenziert sein, d.h. es müssen sowohl ökonomische, als auch politische, kulturelle und gesellschaftliche Aspekte der globalen Integration der Städte berücksichtigt werden, die zudem voneinander unterscheidbar sein müssen. Die explorative Analyse dieser Datengrundlage hat innerhalb eines kohärenten methodischen Rahmens zu erfolgen, der die relationale Natur, die Komplexität und die daraus folgende Grösse des Netzwerkes in den Fokus stellt.

Motivation und Relevanz

Mit dieser Problemstellung wird sowohl ein methodischer als auch ein datenbasierter Ansatz verfolgt. Sie kann jedoch nicht losgelöst vom thematischen Hintergrund betrachtet werden. Zudem sind sowohl der thematische Hintergrund als auch die methodischen Herausforderungen ein inhärenter Teil der geographischen Forschung.

Die geographische Forschung über Städte ist breit und heterogen angelegt. Die Stadt wurde mit unterschiedlichen Perspektiven durchleuchtet und unzählige Ansätze sind dafür entwickelt worden. Pumain und Robic (1996) finden trotz der Breite einen gemeinsamen Nenner aller stadtgeographischen Ansätze: die Stadt wird immer als Teil eines Gebietes oder eines Systems betrachtet und nie in isoliertem Zustand. Die vorliegende Arbeit greift, in dem sie die Stadt als Knoten eines globalen Städtenetzes bzw. als Element eines globalen Systems definiert, diesen gemeinsamen Nenner auf.

Die methodische Problemstellung findet ihre Berechtigung und Relevanz vor allem im Bereich der Geographischen Informationswissenschaft. So beschäftigt sich dieser Forschungsbereich mit der *„Entwicklung und Anwendung von Theorien, Methoden, Technologie und Daten für das Verständnis von geographischen Prozessen, Beziehungen und Mustern. Die Umwandlung von geographischen Daten in Informationen ist zentral für die Geographische Informationswissenschaft“* (übersetzt aus Mark 2003).

Diese Definition greift die wichtigsten Aspekte der vorliegenden Problemstellung auf. Beide fokussieren auf die kohärente und theoriegeleitete Entwicklung und Anwendung von Datengrundlagen und Methoden. Dies erfolgt zudem nicht zum Selbstzweck sondern, um einen (stadt-)geographischen Prozess zu verstehen und dabei neues Wissen über das globale Städtenetz zu generieren.

Wie einführend erwähnt, birgt die Thematik durch ihren relationalen Charakter eine intrinsische Ambivalenz zwischen geographischer und topologischer Nähe und somit im weiteren Sinn zwischen geographischem und topologischem Raum. Durch das Vorherrschen des topologischen Raumes verliert die geographische Lage der Städte an Bedeutung. Eine besondere Nische der Geographischen Informationswissenschaft hat sich der Visualisierung von nicht-räumlichen Daten und somit der Projektion derselben in

den topologischen Raum gewidmet und einen entsprechenden methodischen Rahmen für die Exploration und Analyse entwickelt: das *spatialization framework*. Unter *spatialization* versteht man nach Skupin und Fabrikant (2007) die „*systematic transformation of high-dimensional data sets into lower-dimensional, spatial representations for facilitating data exploration and knowledge construction*“. Mit dieser Perspektive werden hochdimensionale relationale Daten in zwei-dimensionale, räumliche Darstellungen transformiert. Gestützt auf diesen Forschungsgrundsatz lassen sich Städte in ihrer relationalen Dimension auch mit Ansätzen und Methoden der Geographischen Informationswissenschaft erforschen.

Ziele und Forschungsfragen

Die geoinformationswissenschaftliche Motivation gibt zugleich den Anhaltspunkt für das übergeordnete Ziel der Arbeit:

Das Testen neuer Datengrundlagen und die Entwicklung eines theoriegeleiteten methodischen Rahmens für ein fundiertes Verständnis der Stellung der Städte als Knoten im globalen Städtenetz.

Das übergeordnete Ziel teilt sich in drei Unterziele auf und daraus werden drei Forschungsfragen abgeleitet.

Das erste Ziel bezieht sich auf die Erschliessung neuer Datenquellen als direkte Antwort auf das „*dirty little secret*“ (Short et al. 1996). Das *World Wide Web* und insbesondere das Web 2.0 mit sozialen interaktiven und kollaborativen Komponenten fördert die Verbreitung von globalen *crowdsourcing* Projekten. *OpenStreetMap* und Wikipedia sind dafür wohl die bekanntesten Beispiele. Bei einer einfachen Suche des Begriffs *city* in der englischsprachigen Wikipedia ergeben sich knapp eine Million Artikel als Resultat. Das Potential dieser neuen Datenquellen für wissenschaftliche Analysen ist auch im Bereich der Geographischen Informationswissenschaft erkannt. Goodchild (2007) spricht in Bezug zur freiwilligen und kollektiven Produktion von Wissen mit geographischem Bezug von *volunteered geographic information* (VGI). An dieses Potential knüpft die vorliegende Arbeit an. Das erste konkretisierte Forschungsziel dieser Arbeit ist somit:

Die Erschliessung von globalen nutzergenerierten (*crowdsourced*) Datenbanken als neue Datengrundlage für die Erforschung des globalen Städtenetzes.

Die vorliegende Arbeit bedient sich der Wikipedia Enzyklopädie als Datengrundlage. Dies weil sie den in der Problemstellung hervorgehobenen Eigenschaften des globalen Städtenetzes entspricht. Die Forschungsfrage lautet somit:

Ist Wikipedia eine angemessene Datengrundlage für die relationale und funktional differenzierte Erforschung des globalen Städtenetzes und somit eine Antwort auf das „*dirty little secret*“ (Short et al. 1996)?

Das zweite Ziel bezieht sich auf die Entwicklung eines theoriegeleiteten methodischen Rahmens. Einerseits gibt es beim aktuellen Forschungsstand noch nicht ausgeschöpfte Potentiale insbesondere im visuell-analytischen Bereich, andererseits erfordert die Datengrundlage neue methodische Ansätze. Das zweite konkretisierte Ziel ist somit:

Die Entwicklung eines theoriegeleiteten methodischen Rahmens und somit die Erweiterung der aktuellen methodischen Ansätze im betrachteten stadtgeographischen Forschungsbereich.

Dabei sind unterschiedliche Aspekte zu beachten. Erstens postulieren Smith und Timberlake (1995b) aufgrund der relationalen Natur des globalen Städtetetzes einen netzwerkanalytischen Ansatz. Zweitens tangiert die präsentierte Problemstellung wesentliche Elemente der Geographischen Informationswissenschaft, insbesondere der *spatialization*. Drittens ist Wikipedia eine semi-strukturierte semantische Datenbasis und bedingt die Integration semantischer Methoden. In Anbetracht dieser methodischen Aspekte lautet die zweite Forschungsfrage:

Lassen sich mit einer Triangulation von Netzwerkanalyse und -visualisierung, semantischer computerbasierter Analyse und Methoden der *spatialization* die aktuellen methodischen Ansätze im betrachteten stadtgeographischen Forschungsbereich erweitern?

Die zwei methodischen Forschungsfragen werden durch die konkrete Anwendung auf drei Fallstudien über das globale Städtetetz beantwortet. Dabei wird neben den methodischen Zielen auch ein inhaltliches Ziel verfolgt:

Die Generierung eines vertieften Verständnisses über die Stellung der Städte als Knoten im globalen Städtetetz.

Die inhaltlichen Forschungsfragen beziehen sich auf die neuen geographischen Erkenntnisse über die Stellung der Städte als Knoten im globalen Städtetetz, welche man mit dem entworfenen Ansatz gewinnen kann und lauten:

Welche regionalen Strukturen sind im globalen Städtetetz erkennbar? Welche prägen damit die relationale Geographie?

Wie unterscheiden sich die globalisierten Städte hinsichtlich der Zentralität und der Rolle, die sie im globalen Städtetetz übernehmen? Welche Städte können als globale Städte betrachtet werden?

Welche Arten von Beziehungen bzw. welche Funktionen gibt es im globalen Städtetetz? Und wie unterscheiden sich die Netzwerkstruktur und die Zentralität der globalen Städte in unterschiedlichen funktionalen Netzwerken?

Forschungsansatz

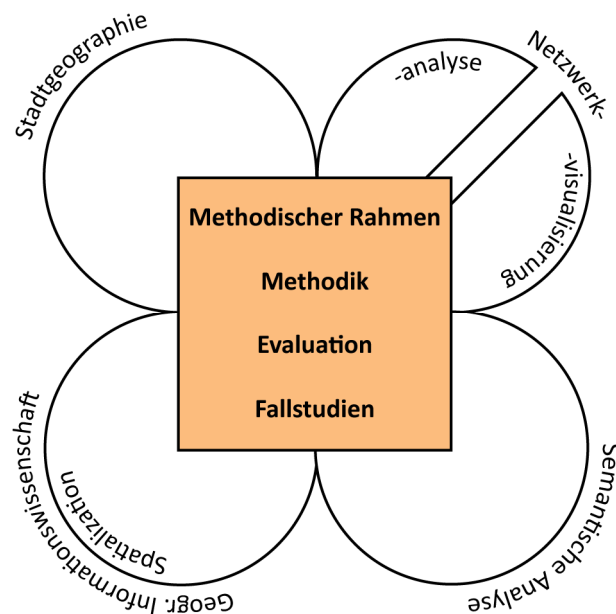
Das gesetzte Ziel umfasst stadtgeographische und geoinformationswissenschaftliche Aspekte. Für die Beantwortung der gestellten Forschungsfragen wird ein interdisziplinärer Ansatz gewählt, um der Komplexität des Forschungsgegenstandes Rechnung zu tragen. Die Stadtgeographie als thematischer Hintergrund und die Geographische Informationswissenschaft, die die Basis für den methodischen Ansatz bietet, werden durch Ansätze und Methoden der Netzwerkanalyse, der -visualisierung und der semantischen Analyse ergänzt. Daraus ergibt sich ein vierstufiger Ansatz (Abbildung 1):

1. **Methodischer Rahmen.** Dieser Rahmen umfasst die Untersuchung und Analyse von Netzwerkdaten im Allgemeinen und ist nicht ausschliesslich auf das Städtetetz ausgerichtet. Er stützt sich auf die *spatialization* und integriert Konzepte

und Ansätze aus der Geographischen Informationswissenschaft und der sozialen Netzwerkanalyse und –visualisierung. Es handelt sich dabei um einen visuell-analytischen Ansatz. Nach der Eruierung der theoretisch-konzeptionellen Grundlage werden zwei konkrete, theoretisch fundierte methodische Arbeitsabläufe hergeleitet.

2. **Methodik.** Gestützt auf den theoretisch-konzeptionellen Rahmen und die methodischen Arbeitsabläufe wird die Methodik für die Untersuchung und Analyse des globalen Städtennetzes konkretisiert. Zudem werden die semistrukturierten Daten aus Wikipedia mittels computerbasierter semantischer Analyse strukturiert und modelliert.
3. **Validierung.** Die Kernpunkte der Methodik werden hinsichtlich ihrer Validität und Sensitivität geprüft.
4. **Fallstudien.** Die entwickelte Methodik wird auf drei Fallstudien angewendet. Die Resultate liefern funktional differenzierte Erkenntnisse über das globale Städtennetz.

Abbildung 1: Forschungsansatz der vorliegenden Arbeit



Aufbau

Der Rest der Arbeit ist wie folgt gegliedert:

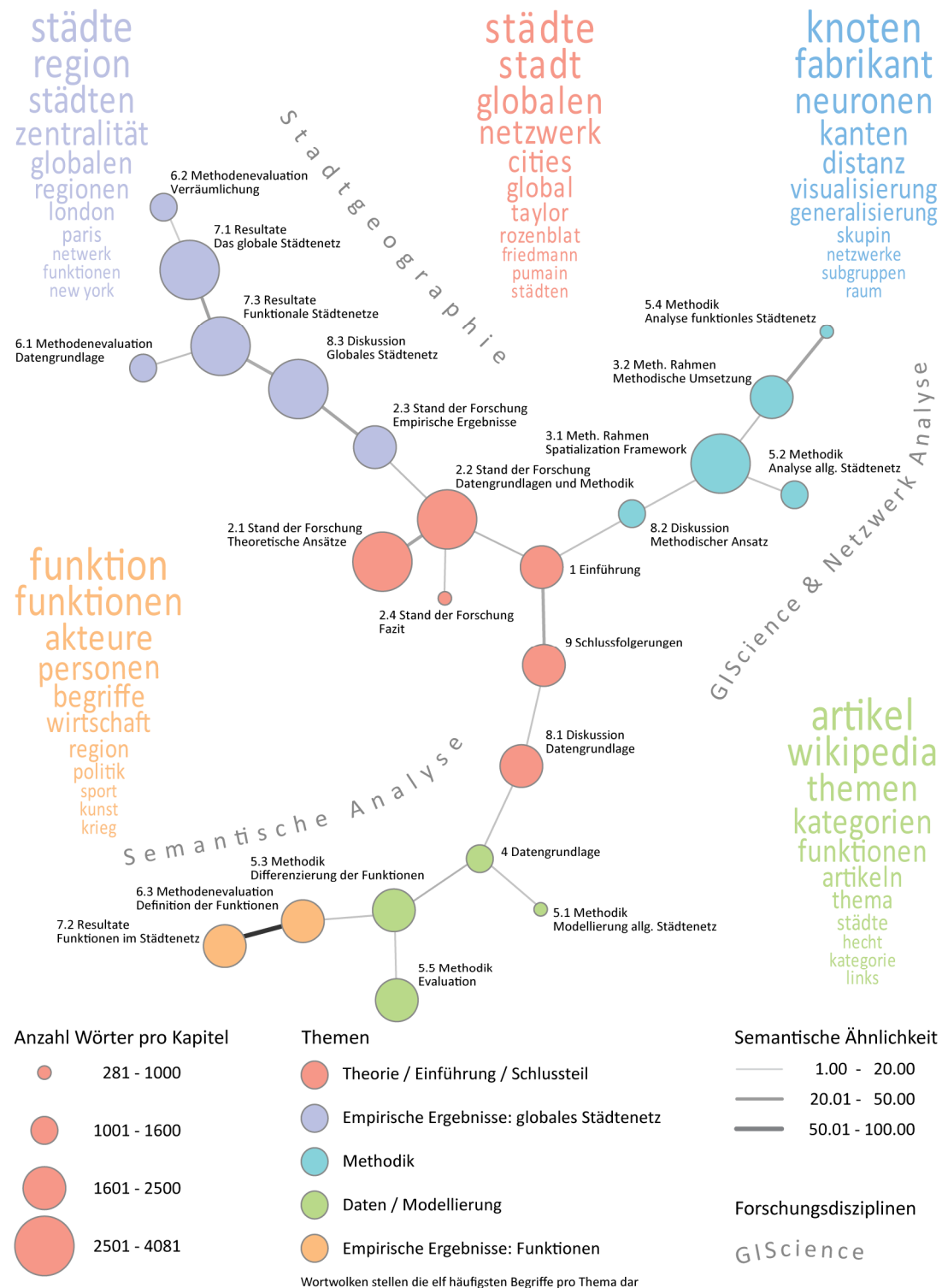
Kapitel 2 fasst den Stand der Forschung im Bereich der *global* bzw. *world cities* zusammen und legt den thematischen Hintergrund fest. Insbesondere werden die wichtigsten Elemente aus der Theorie und den aktuellen methodischen Ansätzen aufgegriffen und daraus die Forschungslücken identifiziert. Kapitel 3 stellt den neu entwickelten methodischen Rahmen vor. Dabei werden die relevanten Aspekte aus dem Forschungsstand der betrachteten methodischen Disziplinen (Geographische Informationswissenschaft, Netzwerkanalyse und semantische Analyse) zu einem kohärenten methodischen Rahmen zusammenführt. Kapitel 4 gibt einen Überblick über die Struktur und die wissen-

schaftliche Erforschung von Wikipedia. Kapitel 5 präsentiert die gewählte Methodik für die Modellierung, Analyse und Validierung des Städtetetzes. Kapitel 6 stellt die Resultate der Methodenevaluation der gewählten Ansätze vor. Kapitel 7 ist den drei Fallstudien gewidmet und legt die wichtigsten thematischen Resultate dar. Kapitel 8 diskutiert die thematischen und methodischen Erkenntnisse. Kapitel 9 fasst die Kernpunkte der Arbeit zusammen und schliesst mit einem Ausblick.

Abbildung 2 fasst die thematischen Schwerpunkte der einzelnen Kapitel graphisch zusammen. Dabei stellen die einzelnen Knoten die Kapitel (zweite Ebene) und die Kanten die semantische Ähnlichkeit in der Form eines *minimum spanning trees* (MST) dar. Die Kapitel gliedern sich in fünf thematischen Schwerpunkte, welche die der Arbeit zugrundeliegenden Disziplinen (Abbildung 1) wieder aufnehmen. Das zentralste Thema (rote Knoten) fasst die allgemeinen Kapitel der Arbeit zusammen: Einführung (Kap. 1), der Schlussteil (Kap. 8.1, 9) und der Grossteil des Stands der Forschung (Kap. 2.1, 2.2, 2.4). Die häufigen Begriffe beziehen sich dabei auf das globale Städtetetz (*Städte / Stadt / globalen / Netzwerk / cities / global / Städten*) oder sind AutorInnen aus dem stadtgeographischen Forschungsbereich (*Taylor / Rozenblat / Friedmann / Pumain*). Ein weiteres Thema (violette Punkte) umfasst die Kapitel aus der Methodenevaluation (Kap. 6.1, 6.2), den Resultaten (Kap. 7.1, 7.3), dem Stand der Forschung (Kap. 2.3) und der Diskussion (Kap. 8.3), welche die empirische Ergebnisse mit Fokus auf das globale Städtetetz darlegen. Die häufigsten Begriffe beziehen sich auf die globalen Städte (*Städte / Städten / London / Paris / New York*) und deren Eigenschaften (*Zentralität / globalen*) und Organisation (*Region / Regionen*).

Der Grossteil der methodischen Kapitel, insbesondere der methodische Rahmen (Kap. 3.1, 3.2), die analytischen Kapitel der Methodik (Kap. 5.2, 5.4) und die Diskussion der Methodik (Kap. 8.2) gehören zu einem weiteren Thema (blaue Punkte). Die häufigsten Begriffe beziehen sich auf die Geographische Informationswissenschaft (*Fabrikant / Neuronen / Distanz / Visualisierung / Generalisierung / Skupin / Raum*) und die Netzwerkanalyse (*Knoten / Kanten / Netzwerke / Subgruppen*). Ein kleineres Thema (grüne Punkte) umfasst die Datengrundlage (Kap. 4) und die Methodik zur Modellierung und Evaluation (Kap. 5.1, 5.3, 5.5). Die häufigsten Begriffe beziehen sich auf Wikipedia (*Artikel / Wikipedia / Kategorien / Artikeln / Kategorie / Links*) oder auf die semantische Analyse (*Themen / Thema / Hecht*). Zwei Kapitel (orange Punkte) umfassen wiederum die empirischen Ergebnisse. Diese beziehen sich jedoch nicht direkt auf das globale Städtetetz, sondern konzentrieren sich auf die Funktionen im Städtetetz. Die wichtigsten Begriffe heben die urbanen Akteure (*Akteure*) hervor, welche den Grundsatz der einzelnen identifizieren Funktionen (*Personen / Wirtschaft / Politik / Sport / Kunst / Krieg*) sind.

Abbildung 2: Thematische Übersicht der Arbeit. Methodik nach Fabrikant und Salvini (2011)



2 Stand der Forschung

Das folgende Kapitel gibt einen Überblick über den Stand der Forschung im Bereich der *global cities* bzw. *world cities*. Der Fokus liegt auf den empirischen Arbeiten im stadtgeographischen Forschungsbereich und weniger auf der zugrundeliegenden Theorie. Im ersten Teil werden die theoretischen Beiträge dargelegt und einander gegenübergestellt. Der zweite Teil befasst sich mit den methodischen Ansätzen. Aus der Übersicht wird insbesondere eine methodische Forschungslücke, die vom Fehlen geeigneter Daten bestimmt wird, hervorgehoben. Im letzten Teil werden die wichtigsten empirischen Erkenntnisse über die Einbettung der Städte im globalen Netzwerk wiedergegeben. Diese Ergebnisse sind zudem eine Vergleichsbasis für die Resultate dieser Arbeit, welche im Kapitel Resultate - Fallstudien dargelegt werden.

2.1 Theoretische Ansätze

Die vorliegende Arbeit stellt die Stadt als Teil eines übergeordneten globalen Ganzen in den Fokus. Für dieses globale Konstrukt werden viele unterschiedliche Bezeichnungen verwendet: *global city network*, *global city system*, *world system of cities*, *world city hierarchy*, usw.¹ Diese semantische Vielfalt ist ein Zeichen für die grosse Anzahl an theoretischen Ansätzen, die über die Zeit entwickelt worden sind (Taylor 2012). Grundsätzlich lassen sich diese Ansätze auf zwei Perspektiven zurückführen. Die systemische Perspektive etabliert sich vor allem in den 1960er bis 1980er Jahren des letzten Jahrhunderts (Pumain und

¹ Eine ausführliche Liste ist in Taylor (2012) zu finden.

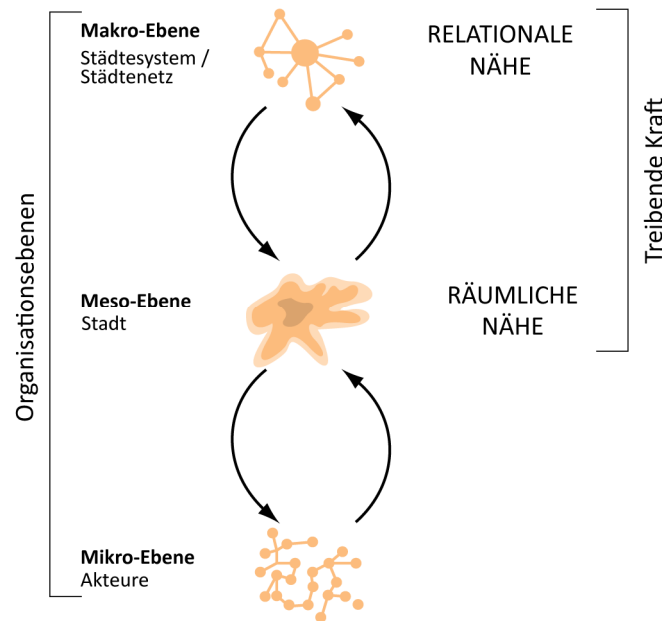
Robic 1996). Eine Stadt ist dabei Teil eines urbanen Systems, in dem die Städte aufgrund ihrer Eigenschaften voneinander abhängig sind. Ein Beispiel dafür ist die vom Prozess der Arbeitsteilung bedingte Spezialisierung der Städte, die zwangsläufig zur Integration der verschiedenen Städte zu einem System führt. Die Wechselbeziehungen sind dabei Ausdruck der unterschiedlichen Eigenschaften und der Abhängigkeit der Städte. Wichtige Begriffe sind hier System und Hierarchie. Die netzwerkanalytische Perspektive hat hingegen ihre Blütezeit ab den 90er Jahren und wird insbesondere von der Globalisierungsdebatte begünstigt (Taylor 2004). Die Stadt kommt dabei aus einem Netzwerk an interstädtischen Beziehungen hervor. Diese Beziehungen sind der Ausdruck der Konzentration von Ressourcen und Macht innerhalb der Städte. Die Art, Intensität und strukturelle Organisation dieser Beziehungen stehen bei der Analyse im Vordergrund (Rozenblat 2004). Kohärent zu diesem Ansatz ist insbesondere der Begriff Netzwerk.

Rozenblat (2004) und Pumain (2006a) integrieren diese beiden Perspektiven zu einem kohärenten Ansatz: „*les réseaux constituent une représentation partielle de l'intégration des villes dans le système urbain*“ (Rozenblat 2004). Mit diesem Ansatz werden einerseits den Beziehungen im urbanen System eine Form und eine Intensität gegeben, andererseits wird die zeitliche Entwicklung der Stadt mit der Netzwerkorganisation und –entwicklung in Beziehung gesetzt (Rozenblat 2004; Pumain 2006b). Bei der Betrachtung der Diskussion und Auslegung des Ansatzes wird mit der Zeit eine Verschiebung hin zu einem stärkeren Akzent auf das Netzwerk ersichtlich. So beziehen sich in Rozenblat (2004) und in Pumain (2006a, 2006b) die gewählten Begriffe eher auf das System. In den aktuelleren Werken (Rozenblat 2010; Pflieger und Rozenblat 2010) ist das Konzept des Netzwerkes allgegenwärtig. Aufgrund dieser Integration wird in dieser Arbeit nicht zwischen Städtesystem und Städtenetzwerk unterschieden, da beide Begriffe das gleiche Phänomen bezeichnen.

Die Integration von Netzwerk und System ist nur ein Aspekt des Ansatzes von Pumain (2006a, 2006b). Pumain (2006a, 2006b) definiert ein urbanes Mehr-Skalen-System. Das oben angesprochene Netzwerk ist eine Ebene davon. Der Kern des theoretischen Rahmens bilden die systemischen Arbeiten von Berry (1964) aus den 60er Jahren des letzten Jahrhunderts. Der Titel seines bekanntesten Artikels ist gleichwohl die synthetischste Definition des Ansatzes: „*cities as systems within systems of cities*“ (Berry 1964). Zwei Skalen werden dabei gleich ersichtlich: erstens die Stadt als System (*cities as systems*) und zweitens das Städtesystem (*system of city*). Darauf aufbauend definiert Pumain (2006a) ein urbanes Drei-Skalen-System, wobei die drei Skalen drei unterschiedlichen Prozessebenen entsprechen, auf denen jeweils spezifische Prozesse stattfinden (Rozenblat 2010). Die Makro-Ebene entspricht dem Städtesystem bzw. dem Städtenetz und somit dem *system of cities* von Berry (1964); die Meso-Ebene bezieht sich auf die Stadt und somit auf das Konzept *cities as systems* von Berry (1964). Die dritte Ebene, die Mikro-Ebene, umfasst die urbanen Akteure. Diese drei Ebenen bilden ein hierarchisches System (Pumain 2006a).

Abbildung 3 fasst das Drei-Skalen-System graphisch zusammen. Im weiteren Verlauf des Kapitels werden die drei Ebenen vorgestellt. Der Fokus liegt dabei auf dem Städtensystem bzw. auf dem Städtenetz. Im Anschluss daran werden drei grundlegende theoretische Beiträge, welche sich mit dem Städtensystem bzw. Städtenetz beschäftigen vertiefter vorgestellt. Diese drei Beiträge sind für die Entwicklung des vorliegenden Ansatzes von massgeblicher Bedeutung.

Abbildung 3: Urbanes Drei-Skalen-System nach Pumain (2006a) und Rozenblat (2010)



Die Mikro-Ebene: die urbanen Akteure

Unter urbanen Akteuren versteht man alle in der Stadt lokalisierten und aktiven sozialen Akteure (Pumain 2006a). Beispiele dafür sind aus dem politischen Bereich die Stadtverwaltung oder die BürgerInnen, aus dem ökonomischen Bereich Unternehmen oder Arbeitskräfte, oder aus dem religiösen Bereich die Kirche. Die Liste der urbanen Akteure könnte unendlich lang sein und zahlreiche Institutionen oder Einzelpersonen umfassen. Die Akteure treffen Entscheidungen (Pumain 2006a) und führen Handlungen durch, die nicht losgelöst vom sozialen Umfeld stattfinden, sondern oft zu Interaktionen und zum Austausch von Ressourcen führen (Rozenblat 2010): z.B. werden Informationen bei interpersonellen Kontakten oder bei gemeinsamen Forschungsprojekten geteilt, Güter werden beim Kauf und Verkauf von Produkten weitergegeben oder Personen wechseln bei Kündigungen und Neuanstellungen von einem Unternehmen zum anderen. Diese Interaktionsnetzwerke, welche sich zwischen den urbanen Akteuren aufspannen, prägen nicht nur die Mikro-, sondern auch die Meso- sowie die Makro-Ebene. Es handelt sich dabei auf allen drei Ebenen um die gleichen Netzwerke zwischen urbanen Akteuren (Rozenblat 2010). Unterschiedlich und ebenen-spezifisch sind jedoch die Prozesse, die diese Netzwerke prägen. Diese Prozesse tragen massgeblich zur Unterscheidung der drei Organisationsebenen bei. Mikro-Prozesse bestimmen dabei den Ursprung des Netzwerkes und dessen Unterhalt (Rozenblat 2010).

Die Meso-Ebene: die Stadt

Auf der Meso-Ebene der Stadt profitieren die Netzwerke zwischen den urbanen Akteuren von der lokalen räumlichen Nähe, welche innerhalb einer Stadt gegeben ist (Rozenblat 2010). Durch die räumlich geförderten Interaktionen zwischen den urbanen Akteuren und deren natürlichen Umwelt (Pumain 2006a) entstehen neue Eigenschaften, die mehr als das einfache Zusammenbringen der Eigenschaften der einzelnen urbanen Akteure sind und somit die Stadt als Entität kennzeichnen (Pumain 2006a). Die Stadt bildet dabei ein soziales und ökonomisches System (Friedmann 1986) oder Netzwerk, welches durch die räumliche Nähe entsteht und sich entwickelt. In ökonomischer Hinsicht spricht man in Bezug auf die durch räumliche Nähe bedingten Externalitäten von Agglomerationseffekten, die dem städtischen System Kohärenz und Dynamik verleihen (Pflieger und Rozenblat 2010). Für eine detaillierte Diskussion der ökonomischen Beziehungen, insbesondere der intra- und interunternehmerischen Interaktionen auf interstädtischer Ebene siehe Rozenblat (2010). Die Stadt besteht jedoch nicht nur aus ökonomischen Interaktionen, sondern aus einer Vielzahl von unterschiedlichen gesellschaftlichen Beziehungen.

Die räumliche Abgrenzung einer Stadt, als Raum in dem Nähe entsteht, stellt eine besondere Herausforderung dar (Pumain 2006a). Wenn man die Stadt als System oder Netzwerk versteht, entspricht sie einem stark verbundenen bzw. integrierten Subsystem oder einer Subgruppe (Friedmann 1986; Wasserman und Faust 2008). Nach dieser nicht-räumlichen Definition kann die Stadt die administrativen Grenzen überschreiten und sich über grössere sogar nicht-zusammenhängende urbane Gebiete ausbreiten (Pumain 2006a). Ein Beispiel für eine mehrpolige Stadt wäre das Ruhrgebiet. Bei solchen mehrpoligen räumlichen Gebilden, bei denen sowohl die räumliche Nähe als auch Beziehungen über grössere Distanzen vorhanden sind, spricht man auch von *global city-regions* oder *mega city-regions* (für eine vertiefte Diskussion siehe Scott et al. 2001; Scott 2001). Die Abgrenzung der Stadt ist eine offene und kontroverse Frage, auf die in den folgenden Kapiteln mit theoretischen Beiträgen von Friedmann (1986) und Sassen (1991) nochmals eingegangen wird.

Die Makro-Ebene: das Städtesystem bzw. Städtenetz

Die intrastädtischen Netzwerke zwischen den urbanen Akteuren auf der Meso-Ebene sind der Grundstein für die Einbettung der Städte ins Städtenetz bzw. die Integration ins Städtesystem (Pflieger und Rozenblat 2010). Bei der übergeordneten Einbettung sind die Städte massgeblich involviert; es sind jedoch wiederum die urbanen Akteure, die die interstädtischen Beziehungen aufspannen (Taylor 2004). Die Prozesse auf der Makro-Ebene sind massgeblich von der relationalen Nähe gesteuert, d.h. neue inter-städtische Beziehungen entwickeln sich mit grösserer Wahrscheinlichkeit zwischen bereits stark integrierten Städten. Die relationale Nähe kann auch durch die Infrastruktur und homogenen Normen und Regeln begünstigt werden (Rozenblat 2010). Während Firmen interstädtische Beziehungen aufbauen, um neue Märkte oder Ressourcen zu erschliessen, handeln Privatpersonen aus familiären oder freundschaftlichen Überlegungen. Beide verfolgen jedoch das gleiche Ziel, d.h. im weiteren Sinn, eine in einer anderen Stadt lo-

kalisierte Ressource zu erschliessen. Dadurch beeinflussen die Meso-Prozesse die Integrationsprozesse auf der Makro-Ebene: eine erhöhte Prozessdynamik auf Meso-Ebene führt zur Entwicklung spezifischer Ressourcen, welche die Attraktivität und die Wettbewerbsfähigkeit und somit die Integration einer Stadt auf der Makro-Ebene erhöht (Rozenblat 2010).

Pumain (2006a) unterscheidet zwischen interstädtischen Interaktionen erster und zweiter Ordnung. Interaktionen erster Ordnung sind Austauschflüsse, seien es Personenbewegungen, Gütertransporte, monetäre Transaktionen oder Austausch an Informationen. Die Interaktionen zweiter Ordnung ergeben sich durch die Interaktionen erster Ordnung. Beispiele dafür sind unterschiedliche politische Bedeutungen, wirtschaftliche Vorteile oder kulturelle Ausstrahlung der Städte. Die hohe politische Bedeutung einer Stadt wäre gemäss diesem Ansatz z.B. die Konsequenz starker Interaktionen politischer Natur erster Ordnung. Die Interaktionen könnten z.B. von einer Stadtregierung, von politisch aktiven Nichtregierungsorganisationen oder von Botschaften gefördert werden.

Smith und Timberlake (1995a) unterscheiden die Interaktionen nach Funktion und Form (Tabelle 1). Die vier Funktionen sind: Wirtschaft, Politik, Kultur und Gesellschaft. Die drei Formen sind Menschen, Güter und Informationen. Tabelle 1 stellt die definierte Typologie grafisch dar und illustriert sie mit Beispielen. Timberlake und Smith (1995a) beziehen sich dabei nur auf die Interaktionen erster Ordnung.

Tabelle 1: Typologie der Beziehungen im Städtesystem nach Smith und Timberlake (1995a)

Form	Menschen	Güter	Informationen
Funktion			
Wirtschaft	Arbeit, Manager, Berater	Kapital, Waren	Geschäftstelefongespräch, Technologietransfer
Politik	Truppen, Botschafter	Militärische Geräte	Politische Abkommen
Kultur	Austauschstudenten, Rockkonzerte	Gemälde, Skulpturen	Spielfilme, Videos
Gesellschaft	Familien, Rotes Kreuz	Entwicklungshilfe	Postkarten

Taylor (2004) definiert in seinen ökonomisch geprägten Analysen vier Akteurtypen, die das Städtensystem gestalten: Dienstleistungsfirmen, Stadtregierungen, Branchenorganisationen und der Nationalstaat. Obwohl diese enge Definition auf einen speziellen Bereich der Wirtschaft zugeschnitten ist, lässt sich eine allgemeine Erkenntnis daraus gewinnen: gewisse Akteure, wie z.B. Unternehmen und Stadtregierungen, agieren direkt und integrieren die Stadt ins Städtensystem, andere, wie Branchenverbände und insbesondere der Nationalstaat, wirken indirekt auf die Interaktionen. Der Staat beeinflusst z.B. mit Zöllen den internationalen Kauf und Verkauf von Gütern und wirkt somit indirekt auf die Interaktion Wirtschaft/Güter nach Smith und Timberlake (1995a).

Aus räumlicher Perspektive ist der Massstab weder eine intrinsische Eigenschaft der Makro-Ebene noch der anderen beiden Ebenen. Die Prozesse auf allen drei Organisati-

onsebenen lassen sich sowohl auf regionaler, als auch auf nationaler, kontinentaler und globaler Ebene analysieren (Rozenblat 2010). Auf Makro-Ebene ist die räumliche Skala wiederum von den urbanen Akteuren bestimmt (Wall und Van der Knaap 2005). Regional agierende Akteure integrieren die Stadt in ein regionales Netzwerk bzw. Städtesystem, gleichwohl bilden sich nationale, kontinentale oder globale Städtesysteme (Pumain 2006a). Bestimmend dafür ist die Reichweite der in der Stadt lokalisierten Akteure (Rozenblat 2010). Bedeutende Städte, z.B. London, sind somit in alle Systeme integriert, andere überschreiten die regionale oder nationale Skala nicht.

Obwohl die Akteure eine entscheidende Rolle sowohl für die Definition der Stadt als System, als auch für ihre Integration ins Städtesystem spielen, ist die Stadt selbst unentbehrlich: sie ist mehr als nur ein Ankerpunkt und ein Standort für die urbanen Akteure, sie koppelt als Knoten eine Mehrzahl von Netzwerken mit unterschiedlichen Funktionen, Formen, sowie räumlichen Reichweiten (Pflieger und Rozenblat 2010).

Die folgenden Unterkapitel beschreiben drei fundamentale Beiträge: „*The world city hypothesis*“ von Friedmann (1986), „*The global cities*“ von Sassen (1991) und Raum der Ströme vs. Raum der Orte von Castells (2001). Diese Beiträge haben den theoretischen Rahmen auf der Makro-Ebene massgeblich beeinflusst. Insbesondere die ersten zwei Theorien legen die Basis für die Unterscheidung zwischen den systemischen und den netzwerk-analytischen Ansätzen. Ansätze, die sich stärker auf Sassen (1991) beziehen sprechen von *global cities*, *global city network* oder *global city system*, Ansätze mit einem engeren Bezug zu Friedmann (1986) sprechen von *world cities*, *world city network* oder *world city system*. Der dritte Beitrag integriert die stadtgeographische Debatte in den grösseren Rahmen der gesellschaftlichen Entwicklung.

The world city hypothesis

Friedmann (1986) publiziert unter dem Titel „*The world city hypothesis*“ ein Referenzwerk (Taylor 2004) für den Forschungsbereich und leistet dabei eine Pionierleistung bei der Integration der Stadt in den Globalisierungsprozess (Soja 2000). Der erste Autor der bereits 70 Jahre vor Friedmann sich eine Weltstadt vorstellt ist jedoch Geddes (1915).

Friedmann (1986) führt den von Hymer (1972) initiierten *economical turn* im Forschungsbereich der globalen Städte weiter und setzt die Entwicklung der Stadt in direkte Abhängigkeit vom globalen Wirtschaftssystem. Der bestimmende Prozess dahinter ist die internationale Arbeitsteilung. Die globalen Städte sind dabei die materielle Form (Friedmann und Wolff 1982) bzw. der räumliche Aspekt (Friedmann 1986) davon. Wenn Friedmann (1995) von Stadt spricht, bezieht er sich auf grosse urbane Räume mit einem stark integrierten ökonomischen und sozialen System, welches die administrativen Grenzen überschreiten kann (Friedmann 1986). Jede Stadt übernimmt aufgrund der internationalen Arbeitsteilung eine bestimmte ökonomische Funktion. Die *world cities* unterscheiden sich von den restlichen Städten, in dem sie ein „*world wide system of control over production and market expansion*“ (Friedmann und Wolff 1982) bilden. Insbesondere die transnationalen Unternehmen üben diese Kontrollfunktion aus (Friedmann und Wolff 1982). Weiter fungieren *world cities* als Hauptsitze von Unternehmen und als Nie-

derlassungsorte für internationale Finanzinstitute und weitere APS (Friedmann und Wolff 1982). Zudem haben die *world cities* auch eine ideologische Kontrollfunktion, indem sie die Produktion und Ausbreitung von Informationen kontrollieren (Friedmann 1986).

Nach Friedmann (1986) führt die Arbeitsteilung zwischen den Städten zur Entstehung von „*relations*“ (Friedmann 1986) mit unterschiedlichen Formen und Intensitäten zwischen den Städten des globalen ökonomischen Systems. Er spricht zudem gleich im ersten Satz seines ersten Werkes (Friedmann und Wolff 1982) von einem „*global network of cities*“. Trotzdem sind die untersuchten Merkmale, insbesondere die Standortverteilung von transnationalen Unternehmen (Friedmann und Wolff 1982) systemischer Natur. Nur in seinem neueren Werk (Friedmann 1995) wird das Netzwerk bzw. die Beschaffenheit der Interaktionen konkretisiert: „*cities serve as centres through which flow money, workers, information, commodities, and other economically relevant variables*“.

Die Struktur des Netzwerks bzw. des Städtesystems ist klar hierarchisch organisiert und ist von einem Zentrum-Peripherie-Muster geprägt. Die Stellung in der Hierarchie wird von der ökonomischen Kontrollfunktion und der unterschiedlichen Integration zwischen den Städten im ökonomischen System bestimmt. (Friedmann 1986)

In räumlicher Hinsicht verbinden die *world cities* die regionalen, nationalen und internationalen Volkswirtschaften zum globalen Wirtschaftssystem. Je nach räumlicher Reichweite des Einflusses lassen sich die *world cities* in globale, multinationale, nationale oder subnationale/regionale Städte einteilen. (Friedmann 1995)

Global cities

Sassen ist mit der Veröffentlichung ihres Buches „*The global city: New York, London, Tokyo*“ (Sassen 1991) zur meist zitierten und einflussreichsten Autorin im Forschungsbereich der globalen Städte geworden (Brenner und Keil 2006).

Sie setzt die ökonomische Tradition im Bereich der globalen Städte fort, legt jedoch den Schwerpunkt nicht auf die transnationalen Unternehmen wie Friedmann (1986), sondern auf „*financial, legal, accounting and advertising firms*“ (Sassen 2002) auch APS genannt. Diese Firmen sind in den *global cities* lokalisiert und übernehmen die Organisation und die Führung der globalen Aktivitäten der transnationalen Unternehmen (Sassen 1996).

Die technologischen Fortschritte und die Globalisierung der Märkte haben die nationale und globale Verteilung der wirtschaftlichen Aktivitäten ermöglicht und gefördert. Diese komplexe globale Architektur fordert Dienstleister, die die Organisation dieses Systems übernehmen und den Austausch an Information zwischen den Systemelementen steuern. Die Aktivitäten werden von den transnationalen Firmen an die APS ausgegliedert, welche sich in den *global cities* konzentrieren. Die gleichzeitige Konzentration und Dispersion der ökonomischen Aktivitäten ist die prägendste Eigenschaft des heutigen ökonomischen Systems und führt zum Aufkommen der *global cities*. In dieser Hinsicht sind

global cities Inhaber wichtiger Ressourcen, die die Globalisierung der Märkte ermöglichen. (Sassen 1995, 2002)

In ihrem ersten Werk reduziert Sassen (1991) das *global cities*-Phänomen auf London, Tokyo und New York. Diese enge Perspektive (Soja 2000) wird in den neueren Publikationen auf mehrere Städte bzw. auf ein „*expanding network*“ erweitert (Sassen 2002).

Auf theoretischer Ebene verfolgt Sassen einen Netzwerk-Ansatz. Sie spricht von einem „*network of cities*“ (Sassen 2002) oder von einem „*global network*“ (Sassen 1996) und hebt die Stärke der *global cities* aus der funktionalen Zentralität in diesem Netzwerk hervor (Derudder 2006). Die empirische Ebene stützt sich bei Sassen jedoch auf nicht-relationale Attribute der Städte (Beaverstock et al. 2000a).

In relationaler Perspektive charakterisiert Sassen (2002) das globale Netzwerk als hierarchisch. In räumlicher Perspektive umfassen die *global cities* ausschliesslich die Innenstadt und nicht wie nach Friedmann (1995) postuliert ein grösseres urbanes Gebiet (Sassen 1991). Weiter stellt Sassen (1991) eine zunehmende Trennung zwischen den *global cities* und dem regionalen sowie dem nationalen Hinterland fest: *global cities* interagieren stärker mit anderen *global cities* als mit dem angrenzenden Umland.

Auf übergeordneter Ebene bricht das globale Netzwerk das Zentrum-Peripherie-Modell auf (Sassen 1994). Obwohl es eine hohe Konzentration an *global cities* in der nördlichen Hemisphäre gibt, kommen auch Städte aus den Entwicklungsländern im Netzwerk vor (Sassen 2002).

Raum der Ströme versus Raum der Orte

„*Der Aufstieg der Netzwerkgesellschaft*“ (Castells 2001) ist Teil eines dreiteiligen soziologischen Werkes (Castells 2001, 2002a; 2003) über das aktuelle Zeitalter. Die darin enthaltenen Thesen werden in „*Communication power*“ (Castells 2009) ergänzt und vertieft.

Die für den Forschungsbereich der globalen Städte entscheidende These betrifft die Organisation der aktuellen Gesellschaft in Form von mehreren Netzwerken. Aufgrund dieser prägenden Organisationsform bezeichnet Castells (2009) die gegenwärtige Gesellschaft als globale Netzwerkgesellschaft oder als Informationsgesellschaft (Castells 2001). Erst der nach den 1950er Jahren eingesetzte technologische Wandel ermöglichte das Aufkommen der gegenwertigen Netzwerkorganisation. Schon frühere Gesellschaften waren in Form von Netzwerken organisiert, diese hatten jedoch stark vertikale Strukturen mit ausgeprägten *command-and-control*-Beziehungen (Castells 2009). Mit dem technologischen Wandel setzt sich die aktuelle Netzwerkorganisation als effizienteste Organisationsform durch. Flexibilität (Fähigkeit sich aufgrund veränderten Bedingungen neu zu organisieren), Skalierbarkeit (Fähigkeit die Grösse anzupassen ohne dabei unter grösseren Störungen zu leiden) und Überlebensfähigkeit (erhöhte Fähigkeit Angriffen und Störungen, dank der dezentralen Organisation, zu widerstehen) sind deren prägende Eigenschaften (Castells 2009). Zudem hat die Netzwerkgesellschaft die Fähigkeit Grenzen zu überwinden und ist somit inhärent global (Castells 2009). Die Netzwerkorganisa-

tion und die damit verbundenen Technologien sind jedoch das Mittel, das Instrument des Globalisierungsprozesses, welcher seinen Ursprung in den ökonomischen, kulturellen, massen-medialen und politischen Dimensionen unserer Gesellschaft findet (Castells 2009).

Trotz der globalen Dimension der Netzwerkgesellschaft, ist sie nicht allumfassend. Gerade in dieser Hinsicht spielt sich eine der wichtigen Machtformen dieser Gesellschaft ab: das Ein- und Ausschiessen, bzw. die Teilnahme und der Ausschluss. Castells (2001) bezeichnet den sozialen Raum, der am Netzwerk teilnimmt, darin eingeschlossen ist, als „Raum der Ströme“. Der ausgeschlossene soziale Raum heisst hingegen „Raum der Orte“. Unter Raum versteht Castells (2001) „*die materielle Verbindung der Gleichzeitigkeit*“. In dieser Perspektive ermöglicht der Raum die Interaktion zwischen zwei oder mehr Personen. Der „Raum der Ströme“ setzt nicht die Nähe im physischen Raum der Interaktionen voraus, sondern die Verbindung der Gleichzeitigkeit erfolgt über die Ströme bzw. über das Netzwerk. Nicht nur der physische Raum, sondern auch die Zeit löst sich im „Raum der Ströme“ auf: die Zeitabfolgen werden durch die Erreichung der Gleichzeitigkeit aufgelöst. Castells (2009) bezeichnet die Netzwerkgesellschaft als zeitlos (*timeless*). Der Raum der Orte ist hingegen physisch abgegrenzt und eigenständig. Seine Bedeutung und Relevanz wird nicht durch die Interaktionen sondern durch die physische Nähe innerhalb dieses Raumes gegeben (Castells 2001). Die Zeit folgt den definierten Rhythmen, wie z.B. in den ausgeprägtesten Fällen dem biologischen Rhythmus (Castells 2009).

Eine für diese Arbeit entscheidende Präzisierung von Castells (2001; 2009) ist die Funktionsabhängigkeit der Netzwerksorganisation bzw. der Teilnahme vs. des Ausschlusses. Die Netzwerkgesellschaft, wie oben erwähnt, umfasst sämtliche Dimensionen der gegenwärtigen Gesellschaft, vom Militär über die Politik bis hin zur Kultur und den Medien (Castells 2002b). Es kann Kapital, Informationen, Bilder oder Töne befördern (Castells 2002b, 2010). Jede Funktion organisiert sich in eigenen Netzwerken, was zu unterschiedlichen und wechselnden Geographien des „Raums der Ströme“ bzw. des „Raums der Orte“ führt.

Die eigentliche Verbindung zwischen der Netzwerkgesellschaft von Castells (2001; 2009) und dem globalen Städtenetz ergibt sich bei der Betrachtung der unterschiedlichen Ebenen, die den „Raum der Ströme“ bilden. Dieser Raum besteht aus drei Ebenen. Die erste Ebene besteht aus der Infrastruktur, die die Ströme über grössere Distanzen ermöglichen (Taylor 2004). Beispiele dafür sind das Internet, Hochgeschwindigkeitstransporte (z.B. Züge und Fluglinien) und Telekommunikationsverbindungen. Die dritte Ebene umfasst die sozialen Räume der führenden „*technokratisch-finanziellen Elite*“ (Castells 2001). Dazu gehören z.B. exklusive Feriendestinationen, Wohnquartiere, Restaurants sowie Freizeiträume (Taylor 2004). Die entscheidende, zweite Ebene entspricht den informationellen Städten, welche nach Castells (2001) ähnlich mit den *global cities* nach Sassen (1991) sind. Diese Städte sind die Knoten des „Raums der Ströme“ und sind somit für diese Arbeit zentral. Die informationellen Städ-

te sind Orte, in denen soziale Akteure die technologische Infrastruktur nutzen (erste Ebene) und damit diese Orte ins Netzwerk integrieren. Je nach betrachteter Funktion ändern sich die teilnehmenden Städte. Je mehr eine Stadt in verschiedenen Netzwerken integriert ist und diese somit räumlich koppelt, desto zentraler und einflussreicher ist eine Stadt. Castells (2001) definiert die Stadt, anders als Sassen (1991) und Friedmann (1986), als einen Prozess und nicht als einen Ort und legt somit den Fokus auf den Integrationsprozess, von dem die Stadt als Knoten im globalen Netzwerk ein Teil ist.

Nachdem in diesem Kapitel die theoretischen Grundlagen skizziert worden sind, beschreibt das folgende Unterkapitel die davon abgeleiteten empirischen Ansätze. Der Fokus liegt dabei auf den unterschiedlichen verwendeten Datengrundlagen und Methoden. Ein besonderes Augenmerk wird den wenigen visuell-analytischen Methoden geschenkt.

2.2 Etablierte Datengrundlagen und methodische Ansätze

Die erste empirische Untersuchung in diesem stadtgeographischen Forschungsbereich stammt von Cohen (1981). Die methodischen Ansätzen haben sich seither weiterentwickelt und verfeinert, dennoch ist die empirische Analyse der theoretisch definierten Städtenetze die „*Achilles Fersé*“ (Beaverstock et al. 2000b) dieses Forschungsbereiches und der Schwerpunkt der vorliegenden Arbeit. Die datenbasierten und methodischen Schwächen werden in der Literatur immer wieder hervorgehoben. So schätzt bereits Korff (1987) in Bezug auf *The world city hypothesis* von Friedmann (1986) die Datengrundlage als inadäquat ein. Die Kritik wird durch den auf theoretischer Ebene verstärkten Fokus auf die Netzwerkorganisation gegenüber der systemischen Organisation (Castells 1996) noch lauter. Short et al. (1996) publizieren im Zusammenhang mit der unzureichenden Datengrundlage einen oft zitierten Artikel: *The dirty little secret of world cities research*. Short et al. (1996) und später auch andere AutorInnen heben folgende kritische Punkte bezüglich der Daten hervor:

- Nach Short et al. (1996), Beaverstock et al. (2000b) und Rozenblat (2004) werden die meisten Datengrundlagen typischerweise auf nationaler Basis gesammelt. Einerseits sammeln die zuständigen statistischen Ämter Daten über alle Städte und Gemeinden innerhalb des eigenen Landes und andererseits werden internationale Daten auf Ebene der Nationen gesammelt und sagen somit nichts über die Einbettung der einzelnen Städte im Städtenetz aus (Knox 1998).
- Nach Short et al. (1996) sowie Smith und Timberlake (2001) beschreiben die Daten z.B. die Anzahl Hauptsitze oder APS-Niederlassungen, nicht die relationale Natur der untersuchten stadtgeographischen Prozesse, sondern nur die Grösse der Stadt. Nach Beaverstock et al. (2000b) ist dies auch die grösste Schwäche der Arbeiten von Sassen (1991) und Friedmann (1986). Der wichtigste Grund dafür liegt in den fehlenden relationalen Datengrundlagen.

- Alderson und Beckfield (2004) sowie Smith und Timberlake (1995a) greifen eine von Short et al. (1996) nicht erwähnte Schwachstelle der Datengrundlage auf: die fehlende Multifunktionalität und die Möglichkeit, diese Funktionen im Datensatz zu differenzieren. Das folgende Zitat von Alderson und Beckfield (2004) illustriert den Kritikpunkt sehr gut: “*To explore and assess [the] statements of the world city hypothesis, one would ideally construct a multirelational network, combining data on economic, political, social, and cultural linkages between cities*”.

Taylor (1997, 2004) hebt zudem hervor, dass die empirische Arbeit methodisch generell sehr schwach sei, und dass Erkenntnisse über Städte vor allem auf Annahmen beruhen. Auch das gewählte Forschungsdesign wird kritisch hinterfragt: die meisten Studien sind Fallstudien, in denen einzelne Städte miteinander verglichen werden (Taylor 2004). Nach Korff (1987) ist jedoch das Einbeziehen einer grossen Anzahl Städte unentbehrlich, um die globalen und mehrskaligen Prozesse zu verstehen.

Im Folgenden werden drei methodische Ansätze vorgestellt, die eine Antwort auf diese Kritik sind. Es handelt sich um die am stärksten etablierten Ansätze für die Modellierung des globalen Städtenetzes. Alle drei Ansätze stützen sich auf internationale und relationale Daten, die als Grundlage für die Exploration und Analyse von grossen globalen Städtenetzen dienen.

Advance producer services als verbindende Akteure

Das *Globalization and World Cities (GaWC) Research Network*, an der Universität von Loughborough, hat sich zum Ziel gesetzt, die Analyse des globalen Städtenetzes basierend auf relationalen Daten zu fördern (GaWC 2012). Sie entwickeln einen methodischen Rahmen, der auf relationalen Prinzipien und sozialen Netzwerkanalysen basiert (Taylor 2001). Theoretisch stützt sich der empirische Ansatz insbesondere auf die Arbeit von Sassen (1991). Damit sind die APS die entscheidenden Akteure in der Entwicklung des globalen Städtenetzes (Taylor 2004). In diesem Unterkapitel werden ausgewählte datentechnische und methodische Elemente dieses Ansatzes vorgestellt.

Das von der GaWC-Forschungsgruppe definierte Netzwerk ist ein bimodales Netzwerk. Die Knoten sind einerseits APS und andererseits Städte (Taylor 2001). Die theoretische Grundlage dafür liegt in der engen Wechselwirkung zwischen dem Handeln der APS und der Integration der Städte im globalen Netzwerk. Modelliert wird das Netzwerk zwischen 316 Städten aufgrund der Intrafirmenbeziehungen zwischen den Niederlassungen von 100 APS. Das eigentliche Städtenetz wird durch die Transformation des bimodalen in ein unimodales Stadt-Stadt-Netzwerk modelliert. Die Interaktionen zwischen den Städten entsprechen der Summe der Intrafirmenbeziehungen. Der Modellierungsansatz beruht auf der Annahme, dass zwischen allen Niederlassungen einer Firma Interaktionen bestehen, die zudem proportional zur Grösse und Wichtigkeit der zwei Niederlassungen sind. Zudem bestimmen diese Interaktionen die interstädtischen Beziehungen.

Die GaWC-Forschungsgruppe definiert drei grundlegende Konzepte für die Analyse des erwähnten Datensatzes: die *global connectivity*, das *hinterworld* und *arenas*:

- Die *global connectivity* jeder Stadt entspricht dem gewichteten *degree* (Summe aller Verbindungen einer Stadt) und wird aus der unimodalen Stadt-Stadt-Matrix berechnet. Je zentraler die Stadt, desto grösser die *global connectivity*. Variationen dieses Kennwertes werden in Taylor et al. (2002b; 2002c) vorgestellt. In Derudder et al. (2010) wird die zeitliche Entwicklung zwischen 2000 und 2008 analysiert.
- Das *hinterworld*-Konzept ist eine Erweiterung des Hinterland-Konzepts (Taylor 2004) und greift das theoretische Postulat von Sassen (1991) auf, gemäss dessen der Bezug der *global cities* zum eigenen Umland zu Gunsten stärkerer Beziehungen zu anderen *global cities* verloren geht. Das *hinterworld* ist somit die Einflusszone einer Stadt in relationaler Perspektive. Zum *hinterworld* einer Stadt gehören alle Städte mit einer überdurchschnittlichen *global connectivity* zu dieser Stadt² (Taylor 2004).
- *Arenas* bestehen aus Städten, die starke, gegenseitige Interaktionen haben. Sie werden mittels der Hauptkomponentenanalyse (Taylor et al. 2002a; Taylor und Walker 2001) oder durch *fuzzy classification* (Derudder et al. 2003; Derudder und Witlox 2004) des bimodalen Stadt-APS-Netzwerks eruiert. Dabei gehören aufgrund der Hauptkomponentenanalyse Städte mit ähnlicher Firmenverteilung zur gleichen *arena*. Mit der *fuzzy classification* werden auch hierarchische Tendenzen erkannt: Städte mit ähnlicher Funktion und hierarchischer Stellung im Netzwerk werden zu einer gemeinsamen *arena* gruppiert. In einem neueren Werk bilden Derudder und Taylor (2005) *arenas* aufgrund von Cliquenanalysen im unimodalen Stadt-Stadt Netzwerk. Zur selben *arena* gehören Städte, die im Netzwerk vollständig miteinander verflochten sind.

In visuell-analytischer Perspektive stellt die GaWC-Forschungsgruppe die Resultate der *global connectivity* und des *hinterworlds* mit Kartogrammen dar (Abbildung 4). Dabei handelt es sich um schematische Darstellungen, die einerseits die grundsätzlichen geographischen Distanzen und Strukturen erhalten und andererseits die Lesbarkeit erhöhen (Taylor 2004). Städte werden als Viereck und der Parameterwert mit Farbton und Helligkeit dargestellt. Abbildung 4 zeigt ein Beispiel eines solchen Kartogramms. Aus den Farbwerten zeigt sich z.B. dass London (LN) und New York (NY) eine sehr hohe und Indianapolis (IN) und Rio de Janeiro (RJ) ein sehr tiefe *global connectivity* aufweisen. Die geographische Distanz zwischen den Städten ist nur ansatzweise erhalten, die geographischen Strukturen z.B. die Kontinente sind trotzdem erkennbar.

² Für mathematische Details siehe Taylor (2004)

Abbildung 4: Kartogramm mit der relativen *global connectivity* gewählter Städte (Taylor et al. 2002b)

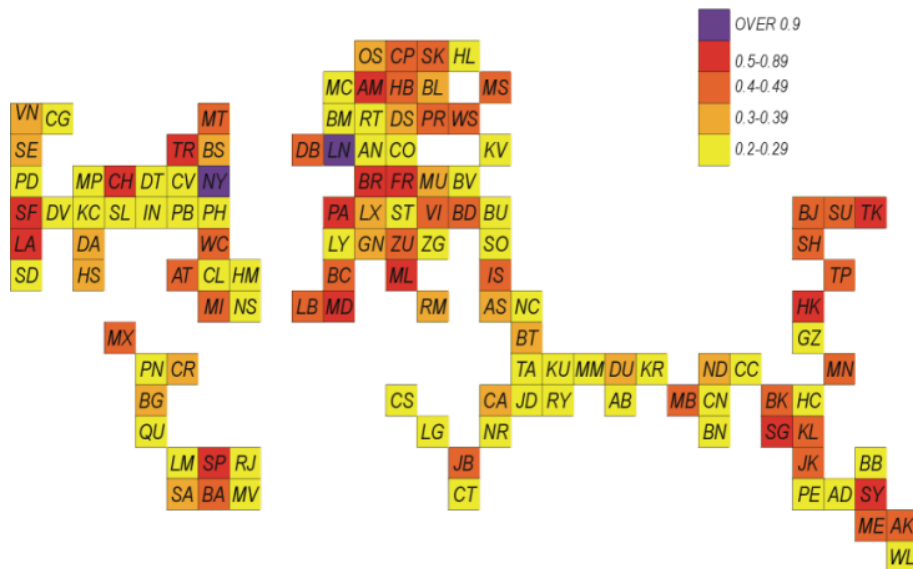
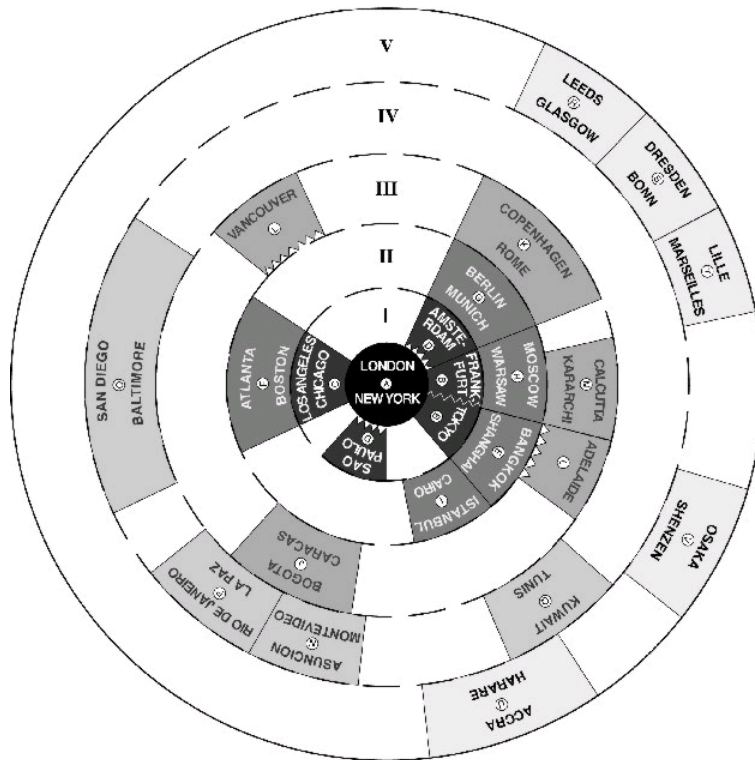


Abbildung 5 ist hingegen ein Beispiel für die graphische Darstellung von *Arenas* (Derudder et al. 2003). Dabei sind die Städte nach ihrer funktionalen und hierarchischen Stellung gruppiert. Die hierarchische Stellung wird aufgrund der Lage in Bezug zum Zentrum des Kreises und der Helligkeit hervorgehoben. Im Zentrum, mit dunkler Farbe hervorgehoben, befinden sich Städte, z.B. London und New York mit einer hohen hierarchischen Stellung. Die funktionale Differenzierung wird durch die sektorale Lage innerhalb der einzelnen Ringe dargestellt. So haben die europäischen Städte, z.B. Amsterdam, Berlin, Rom und Bonn, eine ähnliche Lage innerhalb der jeweiligen Ringe und somit eine ähnliche funktionale Stellung.

Abbildung 5: Urbane *arenas* (Derudder et al. 2003)



Verschiedene Punkte des Ansatzes der GaWC-Forschungsgruppe werden kritisch betrachtet. So wenden Rozenblat und Pumain (2007) ein, dass die Umwandlung von Attributdaten (Firmenniederlassungen) zu relationalen Daten (Intrafirmenbeziehungen bzw. interstädtische Beziehungen) ausschliesslich auf Annahmen beruht. Das Netzwerk wird nicht von relationalen Daten untermauert. Neal (2012) doppelt nach und unterstreichen, dass die Strukturen, welche die vorgelegten Masse hervorheben aufgrund der Datengrundlage, insbesondere der Anzahl Firmen und Städte, vordefiniert sind. Die Strukturen und Muster sind somit datenabhängig und keine valide Abbildung der wahren urbanen Prozesse. Eine Kritik der graphischen Darstellungen ist in der Literatur nicht vorhanden. Aus Perspektive der Kartographie und Geovisualisierung (siehe Kapitel 3) lassen sich diverse Vor- und Nachteile hervorheben. Der Vorteil von beiden Abbildungen ist, dass die Farb-Muster Variablen (Helligkeit und Farbton) den Inhalt salient kodieren. Die zwei graphischen Lösungen unterscheiden sich hingegen hinsichtlich der Lage der Städte: In Abbildung 4 entspricht die Lage ungefähr der geographischen Lage, in Abbildung 5 ist die Lage der Städte aufgrund ihrer Stellung im Netzwerk bestimmt. Der Vorteil von Abbildung 5 ist somit, dass die latenten Strukturen der Daten salienter hervorgehoben werden. Abbildung 4 hat hingegen den Vorteil, dass die geographische Lage der Städte bekannt ist und somit die einzelnen Städte in der Abbildung einfacher zu identifizieren sind. Zwei Nachteile prägen beide Darstellungen: einerseits fehlt eine theoretische Fundierung der gewählten visuell-analytischen Ansätze, andererseits wird das wichtigste Merkmal des Netzwerkes, d.h. die Beziehungen zwischen den Städten nicht graphisch hervorgehoben.

Das nächste Unterkapitel beschreibt einen weiteren Modellierungsansatz und die damit verbundenen Methoden. Dieser Ansatz modelliert das globale Städtenetz aufgrund der Besitzverhältnisse innerhalb der transnationalen Unternehmen und ist eine indirekte Antwort auf die methodischen Schwächen des Modellierungsansatzes der GaWC-Forschungsgruppe.

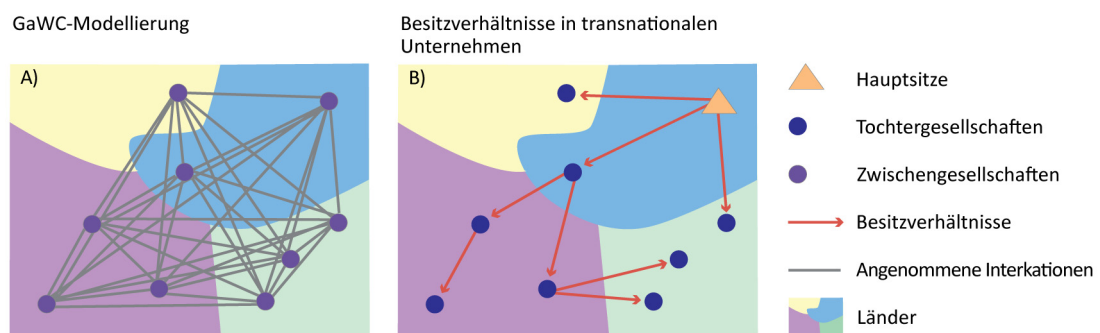
Transnationale Unternehmen als verbindende Akteure

Verschiedene Forschungsgruppen haben das globale Städtenetz aufgrund der Besitzverhältnisse innerhalb der transnationalen Unternehmen modelliert. Theoretisch stützen sie sich schwergewichtig auf den Ansatz von Friedmann (1986) (Derudder 2006). Im folgenden Abschnitt werden insbesondere die Arbeiten von Rozenblat (2004; Rozenblat und Pumain 2007; Rozenblat und Melançon 2009), Alderson und Beckfield (2004; Alderson et al. 2010) und Wall und van der Knaap (2011; Burger et al. 2008) berücksichtigt.

Die Modellierung des Städtenetzes aufgrund der Besitzverhältnisse innerhalb der transnationalen Unternehmen unterscheidet sich vom GaWC-Ansatz nicht nur durch die gewählten Akteure, sondern insbesondere durch die Annahmen, die den Interaktionen zwischen den Firmen zugrunde liegen (Abbildung 6) (Rozenblat und Pumain 2007). Die GaWC-Forschungsgruppe nimmt eine vollständige Interaktion zwischen allen Niederlassungen einer Firma an. Abbildung 6A zeigt das Netzwerk zwischen den Niederlassungen. Die Punktsymbole stellen die Niederlassungen und die Linien die angenommenen Interaktionen dar.

Rozenblat (2004; Rozenblat und Pumain 2007; Rozenblat und Melançon 2009), Alderson und Beckfield (2004; Alderson et al. 2010) und Wall und van der Knaap (2009; Burger et al. 2008) modellieren das Städtenetz nur aufgrund von Besitzverhältnissen innerhalb eines Unternehmens. Es entsteht somit ein gerichtetes Netzwerk zwischen Hauptsitz und Tochtergesellschaften und zwischen Tochtergesellschaften und deren Tochtergesellschaften (Abbildung 6B) (Rozenblat und Pumain 2007). Abbildung 6B zeigt das Besitzverhältnisnetzwerk innerhalb eines transnationalen Unternehmens. Dabei stellen die Punktsymbole unterschiedliche Firmeneinheiten (Hauptsitz, ausländische und intermediäre Tochtergesellschaften) und die gerichteten Linien die Besitzverhältnisse dar.

Abbildung 6: Unterschiedliche Modellierungen des Städtenetzes (Rozenblat und Pumain 2007)



Die empirischen Untersuchungen anhand der Besitzverhältnisse stützen sich mehrheitlich auf eine grössere Anzahl betrachteten Unternehmen. Alderson und Beckfield (2004) stützen ihre Analyse auf 446 Unternehmen der *Global 500* von Fortune für das Jahr 2000. Dabei entsteht ein Netzwerk mit 3'692 Städten. In Alderson et al. (2010) wird die Datengrundlage in zeitlicher und räumlicher Hinsicht erweitert. Wiederum basierend auf den *Global 500* werden die Unternehmensdaten für 1981, 2000 und 2007 für 6'308 Städte zusammengestellt. Rozenblat (2004) untersucht das europäische Städtetz in den Jahren 1990 und 1996. Die Analysen stützen sich auf 300 europäische Unternehmen. Aus geographischer Perspektive werden 225 Agglomerationen mit über 200'000 EinwohnerInnen und weitere 900 bevölkerungsmässig mittelgrosse Städte betrachtet. Nur Wall und van der Knaap (2009) stützen ihre Analysen auf eine beschränkte Datengrundlage: nur 100 Unternehmen aus der *Global 500*-Liste für das Jahr 2005 werden berücksichtigt.

Die Analysen der modellierten Netzwerke haben trotz teilweise signifikanter Unterschiede einen gemeinsamen Nenner: der explizite netzwerkanalytische Ansatz. Die Bedeutung der Städte wird, wie in der vorliegenden Arbeit, mit unterschiedlichen Zentralitätsmassen, wie dem gewichteten *degree*, der nähebasierten Zentralität, der *betweenness* Zentralität oder mit der Rollendefinition nach Guimerà und Amaral (2005a) eruiert (für eine Erklärung der Masse siehe Kapitel 5.2 oder Wasserman und Faust 2008).

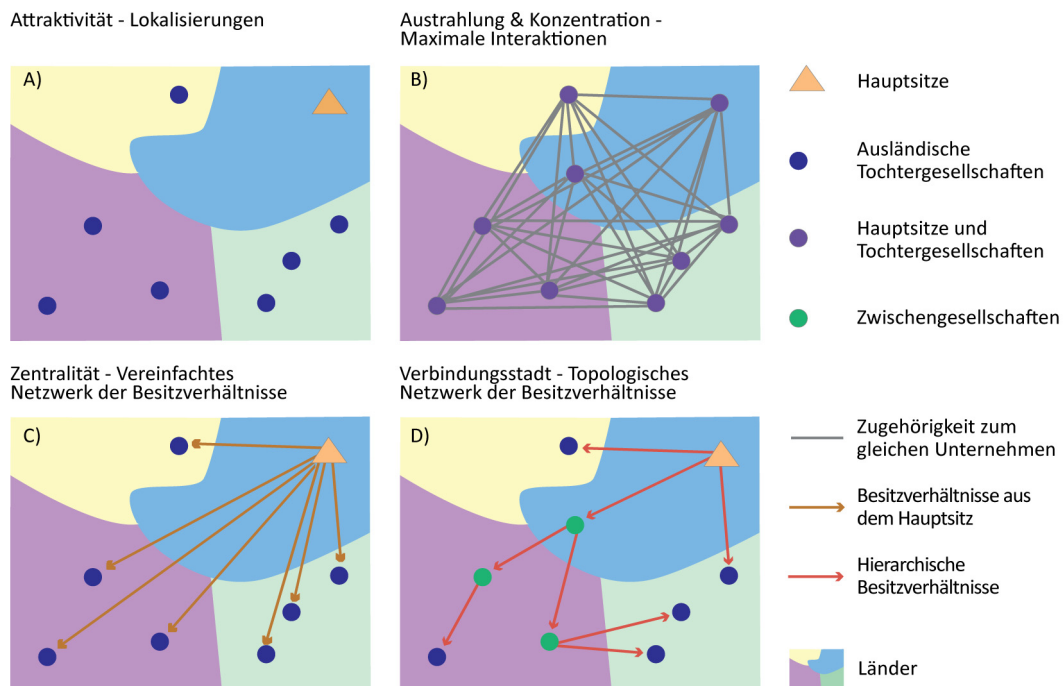
Rozenblat (2004) beschreibt die Stellung der Städte im Netzwerk nicht nur mit Zentralitätsmassen, sondern entwickelt eigene auf das Städtetz massgeschneiderte Masse. Die Modellierung und Analyse des Städtetzes ist dabei sehr differenziert und beruht auf den folgenden vier Annahmen bezüglich der Intrafirmeninteraktionen (Rozenblat 2004):

- Die erste Kennzahl ist die Attraktivität (*attractivités*) der Städte und drückt die Anziehungskraft der Stadt gegenüber den Unternehmen aus. Die Attraktivität entspricht der Summe der Hauptsitze und ausländischen Tochtergesellschaften pro Stadt. Wie Abbildung 7A verdeutlicht wird dabei kein Netzwerk modelliert.
- Die Ausstrahlung (*rayonnement*) und die Konzentration (*Polarisation*) stellen die Einflusszone einer Stadt im Netzwerk bzw. das Potential einer Polbildung dar. Das Netzwerk wird dabei aufgrund der Annahme der vollständigen Intrafirmeninteraktion zwischen allen Tochtergesellschaften modelliert (analog zum GaWC-Ansatz) (Abbildung 7B). Die Ausstrahlung entspricht dem gewichteten *degree* im Gesamtnetzwerk. Für die Konzentration wird das Netzwerk auf die stärkste Interaktion pro Stadt reduziert. Die Konzentration entspricht dem gewichteten *degree* im reduzierten Netzwerk. Konzeptionell entsprechen diese beiden Kennzahlen der *global connectivity* bzw. dem *hinterworld* der GaWC-Forschungsgruppe.
- Die Zentralität (*centralité*) gibt die Kontrollmöglichkeit einer Stadt wieder. Dabei wird das Netzwerk aufgrund vereinfachter Besitzverhältnisse zwischen dem Hauptsitz und den Tochtergesellschaften als gerichtetes Netzwerk modelliert

(Abbildung 7C). Die Zentralität entspricht der Anzahl kontrollierter Tochtergesellschaften.

- Verbindungsstadt (*ville relais*) ist das letzte definierte Konzept. Dabei wird das Netzwerk aufgrund der realen Besitzverhältnisse zwischen Hauptsitz und Tochtergesellschaften aber auch zwischen Tochtergesellschaften modelliert (Abbildung 7D). Verbindungsstädte werden aufgrund ihrer Stellung im Netzwerk und ihrer geographischen Lage in Bezug auf die Ländergrenzen definiert. Somit übernehmen Verbindungsstädte eine besondere Rolle in der Transnationalisierung der Unternehmen.

Abbildung 7: Annahmen für die Modellierung des Städtenetzes (Rozenblat 2004)



Alderson und Beckfield (2004) sowie Wall und van der Knaap (2009) stützen ihre Analysen ausschliesslich auf diese Modellierungsalternative (Abbildung 7D) und berechnen dabei vier netzwerkanalytische Masse, die die unterschiedlichen Aspekte der Macht und des Prestiges der Städte im Netzwerk abbilden.

- Der *outdegree* (siehe Kapitel 5.2 oder Wasserman und Faust 2008) entspricht der gewichteten Summe aller Besitzverhältnisse der in einer Stadt angesiedelten Unternehmen. Je grösser der *outdegree* ist, desto stärker ist die ökonomische Kontrollfunktion einer Stadt.
- Die nähebasierte Zentralität (siehe Kapitel 5.2 oder Wasserman und Faust 2008) stützt sich auf die Länge der Pfade im Besitzverhältnisnetzwerk und hebt Städte hervor, welche andere Städte im Netzwerk gut erreichen können.
- Die *betweenness* Zentralität (siehe Kapitel 5.2 oder Wasserman und Faust 2008) hebt die Brückenfunktion einer Stadt hervor. Nach Rozenblat (2010) wird je-

doch die *betweenness* Zentralität der Städte aufgrund der Eigenschaften der Datengrundlage überschätzt.

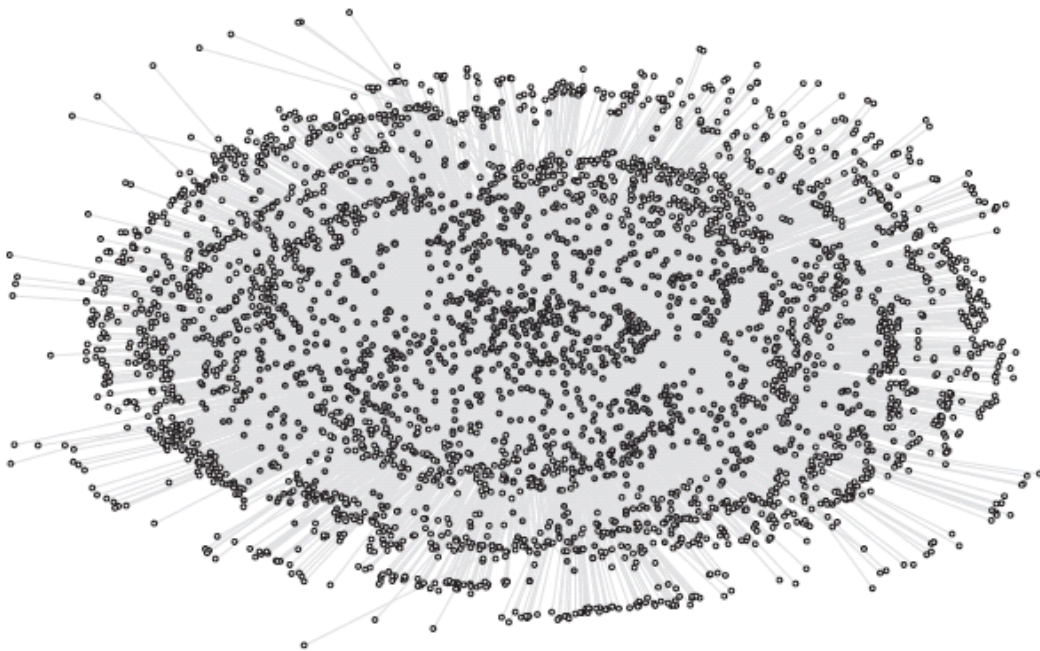
- Der *indegree* (siehe Wasserman und Faust 2008) entspricht wie der *outdegree* der gewichteten Summe der Besitzverhältnisse, jedoch aus der Perspektive der zum Unternehmen gehörenden Tochtergesellschaften in einer Stadt. Je grösser der *indegree* ist, desto grösser das Prestige einer Stadt.

Da in der vorliegenden Arbeit das Städtetz ungerichtet ist, wird nicht zwischen *in*- und *outdegree* unterschieden. Die Eigenschaften der Daten dieser Arbeit beeinträchtigen die Berechnung der *betweenness* Zentralität hingegen nicht.

Neben der Stellung der einzelnen Städte im Netzwerk untersuchen Rozenblat und Melançon (2009) die regionale Organisation im Städtetz mit einer breiten Palette an netzwerkanalytischen Clustermethoden, welche eng verflochtene Gruppen von Städten erkennen. Alderson und Beckfield (2004) hingegen bilden Gruppen von Städten mit ähnlichen Beziehungsmustern mit der traditionellen Methode der Blockmodelle (für Details siehe Wasserman und Faust 2008).

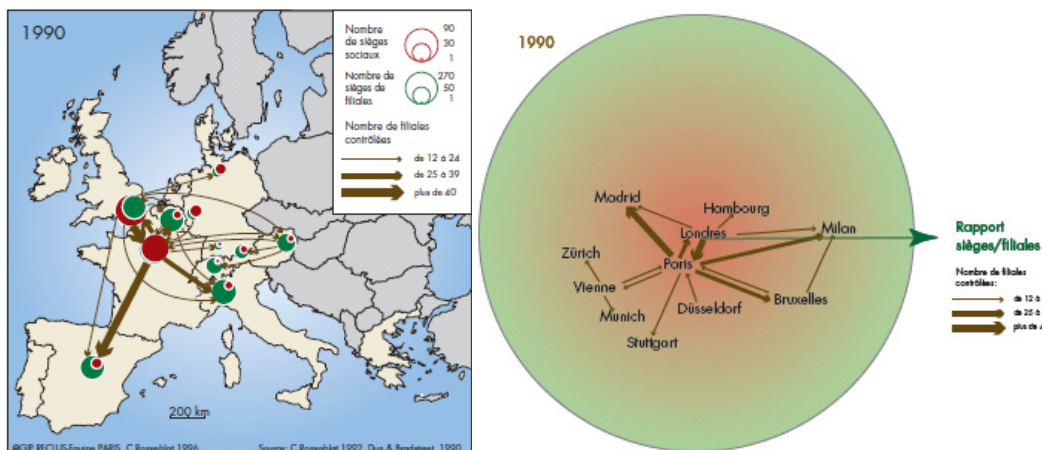
Der netzwerkanalytische Akzent der Forschungsgruppen, die sich mit den Besitzverhältnissen innerhalb der Unternehmen beschäftigen, wird auch durch die verwendete Software bestätigt. Alderson und Beckfield (2004) setzten für die Netzwerkanalysen UCINET (Borgatti et al. 2002) ein. Diese Software umfasst viele der klassischen netzwerkanalytischen Masse und hat eine Rechenkapazität für Netzwerke mit maximal tausenden Knoten. Die Möglichkeiten zur Visualisierung von Netzwerken sind in der Software noch wenig ausgeschöpft. Dies zeigt sich z.B. an der Netzwerkvisualisierung in Abbildung 8. Diese ist mit UCINET generiert und zeigt das globale Städtetzsystem (Alderson et al. 2010), wobei die Punktsymbole die Städte und die Liniensymbole die Beziehungen zwischen denselben darstellen. Obwohl alle Städte und alle interstädtischen Beziehungen gleichzeitig dargestellt werden, ist es sehr schwierig sich ein klares Bild über die latenten Strukturen zu machen. Wegen der fehlenden Beschriftung ist es nicht möglich die Städte zu unterscheiden und die grosse Anzahl Linien überlädt die Darstellung, was letztlich dazu führt, dass die einzelnen Beziehungen nicht mehr voneinander unterscheidbar sind.

Abbildung 8: Globales Städtesystem im Jahr 2007 nach Alderson et al. (2010)



Rozenblat und ihre Forschungsgruppe (2010; Rozenblat und Melançon 2009) setzen in neueren Arbeiten die Software Tulip (LaBRI 2010) als visuell-analytisches Werkzeug ein. Tulip (LaBRI 2010) integriert neben einer Vielzahl von netzwerkanalytischen Massen aus unterschiedlichen Disziplinen (Soziologie, Informatik, Physik, usw.) auch Algorithmen zur Visualisierung von Netzwerkdaten. Bereits in früheren Arbeiten legt Rozenblat (2004) besonderen Wert auf visuell-analytische Methoden und insbesondere auf kartographische Darstellungen. Ein Beispiel dafür wird in Abbildung 9 erläutert. Abbildung 9a zeigt die betrachteten europäischen Städte im geographischen Raum. In Abbildung 9b hingegen sind die Städte aufgrund ihrer Eigenschaften in einem topologischen Informationsraum platziert. Die grünen und roten Kreisdiagramme in Abbildung 9 stellen die Städte dar, wobei die Grösse der grünen Kreise die Anzahl Tochtergesellschaften pro Stadt und die Grösse der roten Kreise die Anzahl Hauptsitze pro Stadt darstellen. Die gerichteten Linien bezeichnen die Anzahl kontrollierter Tochtergesellschaften. In Abbildung 9b werden die gleichen Informationen im topologischen Informationsraum dargestellt. Die Lage der Städte gibt dabei das Verhältnis zwischen Anzahl Hauptsitze und Tochtergesellschaften wieder: zentrale Städte haben einen höheren Anteil an den Hauptsitzen. Die zwei Farbtöne werden redundant zur Lage verwendet und sind gleich wie in Abbildung 9a eingesetzt. Die gerichteten Linien entsprechen den Kontrollbeziehungen aus Abbildung 9a

Abbildung 9: Zentralität der europäischen Städte als a) kartographische (links) und b) schematische Darstellung (rechts) (Rozenblat 2004)



Bei einem Vergleich der Abbildung 4 und Abbildung 8 mit der kartographischen und schematischen Darstellungen von Rozenblat (2004) in Abbildung 9 zeigt sich der ausgeprägte Gewinn an Salienz bei der konsequenten Integration von kartographischen Designprinzipien bei der Visualisierung des Städtennetzes. In die gleiche Richtung geht ein Beitrag von Thierstein und Förster (2008), die für eine verstärkte und konsequentere Verwendung von Visualisierungen, im Spezifischen von Netzwerkvisualisierungen, als Instrument für die Untersuchung von komplexen Daten, wie das globale Städtennetz plädieren. An den kartographisch inspirierten Darstellungen von Rozenblat (2004) und der Aufforderung von Thierstein und Förster (2008) knüpft die vorliegende Arbeit an und erweitert das methodische Feld aus und gliedert Methoden aus der *spatialization* im Spezifischen und der Geographischen Informationswissenschaft im Allgemeinen ein. Zudem werden die verwendeten Designprinzipien und Methoden mit theoretischen und empirischen Erkenntnissen untermauert und zu einem kohärenten methodischen Rahmen integriert (siehe Kapitel 3).

Eine weitere Stärke aller Arbeiten, welche auf den intraunternehmerischen Besitzverhältnissen basieren, ist der starke Einbezug der Netzwerkanalyse. Bereits in den 90er Jahren plädieren Smith und Timberlake (1995a) für einen netzwerkanalytischen Ansatz für die Untersuchung des globalen Städtennetzes. Die netzwerkanalytischen Methoden würden dem relationalen Charakter der betrachteten urbanen Prozesse Rechnung tragen. Dieser methodische Grundsatz wird auch in der vorliegenden Arbeit aufgenommen und weiterentwickelt (siehe Kapitel 3 und 5).

Die beiden bisher betrachteten Ansätze kommen den ersten zwei der oben erwähnten Kritikpunkten bezüglich der Datenlage nach: erstens, beide Datensätze sind global und ermöglichen somit die Untersuchung der Integration der Städte auf globaler und nicht nur nationaler Ebene. Zweitens, die Besitzverhältnisse und in Ansätzen auch die Intra-firmenbeziehungen der GaWC-Forschungsgruppe sind relationale Daten. Ein Kritikpunkt bleibt jedoch offen: beide Datensätze beinhalten nur ökonomische Beziehungen zwischen den Städten. Politische, kulturelle und soziale Interaktionen (Tabelle 1), welche nach Smith und Timberlake (1995a) sowie nach Alderson und Beckfield (2004) gleich

wie die ökonomischen Interaktionen die Städte prägen und Städtetnetze aufspannen, werden weder erkannt noch untersucht.

Das folgende Unterkapitel stellt die wenigen funktionsübergreifenden Datensätze vor und zeigt, warum die Erschließung von weiteren Datenquellen trotzdem notwendig ist.

Multifunktionale Interaktionen zwischen den Städten

Einer der ersten Versuche überhaupt, das globale Städtetnetz zu modellieren, basiert auf Flugpassagierbewegungsdaten (Smith und Timberlake 1995a). Inhaltliche Überlegungen rechtfertigen die Flugpassagierbewegungen als Datengrundlage. Da das Flugzeug ein effizientes Verkehrsmittel ist, um weite Distanzen zu überwinden, sind internationale Flughäfen eine zentrale Komponente der Infrastruktur einer Weltstadt und das Fliegen eine der wichtigsten Erscheinungen der Integration der Städte in ein globales Städtesystem (Smith und Timberlake 2001). Flugpassagierbewegungen umfassen nicht nur ökonomische Interaktionen, sondern sind eine Näherungsvariable für die multifunktionale Integration der Städte im globalen Netzwerk. Von den drei von Smith und Timberlake (2001) postulierten Formen der Interaktionen (Tabelle 1), werden jedoch nur Personen berücksichtigt. Güter und Informationen werden von den Flugpassagierbewegungen nicht abgebildet. Das Flugpassagierbewegungsnetzwerk wird von Smith und Timberlake (1995b, 2001) zur Beantwortung von stadtgeographischen Fragestellungen untersucht. Zudem wird dieser Datensatz für die Evaluation von neuen netzwerkanalytischen Methoden verwendet. Beispiele dafür sind Rozenblat et al. (2010), Guimerà et al. (2005) und Salvini et al. (2011).

Obwohl Flugpassagierdaten explizit global, relational und ansatzweise auch multifunktional sind, haben sie verschiedene Nachteile in Bezug auf die Untersuchung des globalen Städtetnetzes. Der erste Nachteil ist konzeptioneller Natur. Die Infrastruktur, im spezifischen die internationalen Flughäfen, ist nach dem theoretischen Rahmen von Castells (2001) zwar eine Ebene vom „Raum der Ströme“, jedoch nicht die für die Entwicklung und Integration der Städte entscheidende zweite Ebene. Mit der Analyse der Flugpassagierdaten bleiben somit entscheidende Aspekte des globalen Städtetnetzes unberücksichtigt (Taylor 2001). Der zweite Nachteil ist durch zwei Verzerrungen in den Flugpassagierdaten gegeben. Da Touristen nicht von anderen Passagieren getrennt werden, wird die Zentralität von touristischen Städten überschätzt (Gnos 2010; Derudder und Witlox 2008). Da sich zudem die Angaben zu den Flugpassagierbewegungen auf Start- und Landeflughafen und nicht auf Reisestart und –destination beziehen, spiegelt das Netzwerk auch die Flugstrategien der Fluggesellschaften und nicht nur die urbanen Prozesse wider (Derudder und Witlox 2008). Der letzte und wichtigste Nachteil der Flugbewegungsdaten ist die Undifferenziertheit des Netzwerkes. Obwohl das Flugpassagierbewegungsnetzwerk nicht nur eine Annäherungsvariable für ökonomische, sondern auch für politische, soziale und kulturelle Interaktionen zwischen den Städten ist, ist dessen Beitrag für die Untersuchung des multifunktionalen Städtetnetzes gering. Es ist nämlich nicht möglich die vier erwähnten Funktionen (Tabelle 1) voneinander zu unter-

scheiden und somit die funktionsspezifische Einbettung der einzelnen Städte im globalen Netzwerk zu analysieren.

Zwei neuere Studien der GaWC-Forschungsgruppe sind zum jetzigen Zeitpunkt die einzigen Versuche ein multifunktionales, differenzierbares globales Städtenetz zu modellieren. Taylor (2005a, 2005b) weitet dabei die Datengrundlage der GaWC-Forschungsgruppe aus und betrachtet neben den APS andere global tätige Akteure (Tabelle 2). Die Modellierung der Beziehungen folgt mehrheitlich demselben Ansatz wie oben beschrieben. Die ökonomische Zentralität einer Stadt wird mit den APS und wichtigen multinationalen Unternehmen modelliert. Die kulturelle Funktion einer Stadt stützt sich auf Daten über Medien-, Architektur- und Ingenieurfirmen. Die politische Zentralität einer Stadt beruht auf Niederlassungen von Botschaften, UN-Agenturen und städtischen Organisationen. Der gesellschaftliche Aspekt schliesst die Niederlassungen von *non governmental organizations* (NGOs) und die Co-Autorenschaft in wissenschaftlichen Artikeln mit ein. Dabei wird der Standort der Forschungsanstalt, in der die AutorInnen arbeiten, als räumliche Verankerung verwendet. Wichmann Matthiessen et al. (2010) folgen dem gleichen Ansatz und modellieren das globale Städtenetz des Wissens basierend auf der Co-Autorenschaft in wissenschaftlichen Artikeln.

Tabelle 2: Datensätze für die Modellierung multifunktionaler Netzwerke (Taylor 2005a)

Funktion	Datensatz
Wirtschaft	100 APS
	446 multinationale Unternehmen aus <i>Fortune's Global 500</i>
Kultur	33 globale Medienkonglomerate
	44 globale Architektur- und Ingenieurfirmen
Politik*	76 UN-Agenturen
	195 nationale Botschaften
	3 städtische globale Organisationen
Gesellschaft	63 NGOs
	2'768'615 wissenschaftliche Artikel

* In Taylor (2005b) wird die politische Funktion anhand von 133 nationalen Botschaften, 78 UN Agenturen und 74 NGOs modelliert.

Diese Beiträge von Taylor (2005a, 2005b) sind ein gelungener Versuch die Multifunktionalität im Städtenetz zu untersuchen, weisen trotzdem gewisse Schwächen auf. Erstens, gilt auch für diesen Ansatz, wie für den etablierten ökonomischen Ansatz der GaWC-Forschungsgruppe, die Kritik von Neal (2012) bezüglich der Validität und des strukturellen Determinismus. Zweitens, fehlt eine vertiefte Begründung der Operationalisierung der vier Funktionen. Die vorliegende Arbeit schliesst sich der multifunktionalen Modellierung des Städtenetzes an und ist ein Versuch, die multifunktionale und valide Datenquelle Wikipedia zu testen und zu erschliessen.

Dieses Unterkapitel hat sich bisher ausschliesslich mit der Methodik im betrachteten stadtgeographischen Forschungsbereich befasst. Das folgende Unterkapitel fasst hingegen die inhaltlichen Ergebnisse zusammen.

2.3 Überblick der empirischen Ergebnisse

Die empirischen Erkenntnisse über Städte und ihrer globalen Integration sind in der Literatur vielfältig dokumentiert. Dieses Kapitel versucht, die wichtigsten empirischen Erkenntnisse aus ausgewählten Studien³, die das vollständige globale Städtetz untersuchen, zusammenzufassen. Die hauptsächlich betrachteten Studien und deren Eigenschaften sind in Tabelle 3 zusammengefasst. Weitere Studien werden bei Bedarf ergänzend beigezogen. Für jede Studie sind der theoretische Ansatz, die Datengrundlage, die Zentralitätsmasse und weitere Informationen über die Taxonomie der globalen Städte angegeben. Diese wurden im vorgängigen Unterkapitel erklärt und diskutiert. Alle betrachteten Zentralitätsmasse, mit Ausnahme von Friedmann, entsprechen einem gewichteten *degree*, d.h. der gewichteten Summe aller Beziehung einer Stadt zu allen anderen im Netzwerk. Wall und van der Knaap (2011) sowie Alderson und Beckfield (2004; Alderson et al. 2010) unterschieden zudem zwischen ausgehenden (*outdegree*, Kontrollfunktion) und einkommenden (*indegree*, Prestige) Beziehungen einer Stadt. Da das Städtetz der GaWC-Forschungsgruppe hingegen ungerichtet ist, wird nicht zwischen *outdegree* und *indegree* unterschieden.

³ Alle in diesem Unterkapitel nicht weiter präzisierten Referenzen stammen aus den Studien in Tabelle 3.

Tabelle 3: Ansätze zur Untersuchung des globalen Städtenetzes

Forschung	Theoretischer Ansatz	Datengrundlage	Zentralitäts-masse	Taxonomie	Bezugs-jahr	Anzahl Städte
GaWC (2009b, 2009a, 2010, 2011)	Global cities	APS: Intrafirmen-beziehungen	Global connectivity	α -, β -, γ -Städte	2000 2004 2008 2010	> 100
Wall und v.d. Knaap (2009)		APS: Besitz-verhältnisse	Outdegree (OD) Indegree (ID)	Rangierung	2005	30
Alderson und Beck-field (2004; Alderson et al. 2010)	World cities	Transnationale Unternehmen: Besitz-verhältnisse			2000	50
					2007	15
Friedmann (1995)		Konzeptionelle Definition	Konzeptionelle Taxonomie	Global financial articulation (Gfa) Multinational art. (Mna) Important national art. (Ina) regional art. (Ra)	1995	30

Bei der Literaturdiskussion werden in einem ersten Teil insbesondere die folgenden zwei Fragen beantwortet: Welche Verteilungsmuster weisen die Städte in geographischer Hinsicht auf? Welche Städte bilden den Kern des globalen Netzwerkes?

Tabelle 4 bis Tabelle 7 geben eine Übersicht über ausgewählte globale Städte. Daraus lassen sich vier global gültige Trends ableiten. Erstens, die Unterschiede zwischen *global cities* und *world cities* sind nicht besonders ausgeprägt. Dies wird beim umfassenden Vergleich, aber vor allem beim Vergleich zwischen *global cities* und *world cities* in der Studie von Wall und van der Knaap (2009), in der ein Datensatz aus beiden theoretischen Perspektiven untersucht wird, ersichtlich. Zweitens, sind die Zentralitätswerte tendenziell zeitlich konstant. Drittens, Nordamerika, Europa und Asien profilieren sich als Regionen mit hoher Ballung an globalen Städten. Wobei sich die asiatischen globalen Städte fast ausschliesslich im ost- und südostasiatischen Raum konzentrieren. Viertens, fällt das Verhältnis zwischen *outdegree* und *indegree* in Asien, Ozeanien, Süd-, Zentralamerika und Afrika zugunsten des *indegrees* und in Europa und teilweise in Nordamerika zugunsten des *outdegrees* aus. Das heisst, dass die Zentralität im Städtenetz der nordamerikanischen und europäischen Städte vor allem aufgrund ihrer Kontrollfunktion gegeben ist, für die asiatischen, ozeanischen, süd-, zentralamerikanischen und afrikanischen Städte ist die Zentralität massgeblich vom Prestige bestimmt. Die Ballung an Städten in den erwähnten Regionen lässt sich auch mengenmässig teilweise spezifizieren: etwa ein Drittel bis

die Hälfte der globalen Städte befinden sich in Europa, ein guter Drittel in Nordamerika und etwa 10-20% in Ost- und Südostasien. Die Verteilung wird wenig vom gewählten theoretischen Ansatz (*global cities* vs. *world cities*) beeinflusst. Grösser ist der Einfluss der Anzahl der erwähnten Städte und des betrachteten Zentralitätsmasses: wenn man die zentralsten Städte des Netzwerks oder nur die Kontrollfunktion betrachtet ist die Konzentration in Europa und Nordamerika ausgeprägter als wenn mehr Städte oder das Prestige betrachtet wird.

Tabelle 4 zeigt eine Übersicht der Zentralitätsmasse für ausgewählte europäische Städte. Die zentralste Stellung haben je nach Ansatz London oder Paris inne. Wobei London nach der GaWC-Forschungsgruppe und Friedmann sowie aufgrund des *indegrees* und Paris bei der Betrachtung des *outdegrees* herausragt. Diese zwei Städte unterscheiden sich hinsichtlich der konstant sehr hohen Zentralität von den anderen Städten. Starke Evidenz ergibt sich auch für die hohe Zentralität von Frankfurt, Amsterdam, Zürich, Brüssel, Madrid und Mailand. Die letzten zwei südeuropäischen Städte sowie Brüssel profilieren sich, anders als die erwähnten nordeuropäischen Städte wegen des höheren *indegrees*. München und Düsseldorf haben eine vergleichbare Kontrollfunktion (*outdegree*) wie Frankfurt und Zürich, jedoch ein tieferes Prestige (*indegree*), sowie eine tiefere *global connectivity* und werden nach Friedmann nur zu den *regional articulation cities* gezählt. Stockholm, Barcelona, Wien sowie Dublin haben eine geringe Zentralität, die insbesondere aufgrund des Prestiges gegeben ist. Die Kontrollfunktion ist gering.

Tabelle 4: *Global cities und world cities in Europa*⁴

	<i>Global cities</i>				<i>World cities</i>								
Stadt	GaWC-Forschungsgruppe				Wall und v.d. Knaap				Alderson und Beckfield				Friedmann
	<i>Global Connectivity</i>				OD	ID	OD	ID	OD	ID	OD	ID	
	2000	2004	2008	2010	2005				2000		2007		1995
London	α++	α++	α++	α++	3	2	4	2	4	2	4	2	Gfa
Paris	α+	α+	α+	α+	2	5	2	5	3	3	2	3	Ina
Frankfurt	α	α	α-	α	5	17	16	19	11	30			Mna
Amsterdam	α	α	α-	α	7	8	9	16	6	8	6	14	Mna
Zürich	α-	α	α-	α-	4	10	5	20	7	34	8		Ina
Brüssel	α	α	α	α	9	12	17	7		7		9	
Madrid	α	α	α	α		13		8	45	12		15	Ina
Mailand	α	α	α+	α		7		9		14			Ra
München	β	α+	β-	α-	13	26	7	28	8	20	5		Ra
Düsseldorf	β	β+	β-	β+	12		6		5	27	9		Ra
Stockholm	α-	α-	α-	β+					14	26			
Barcelona	α-	β+	β+	α-				21		40			Mna
Wien	β+	β+	α-	α-				17		35			
Dublin	α-	α-	α-	α-		18		14		43			
Berlin	β	β+	β	β+	27								
Rom	β	β	α-	β+					30				
Moskau	β+	β+	α	α									
Istanbul	β+	β	α-	α-									
Warschau	β+	α-	α-	α-									
St. Petersburg			γ-	γ									

Weitere europäische Hauptstädte, wie Berlin, Rom oder in Osteuropa Moskau, Warschau und Istanbul haben weder als *global cities* noch als *world cities* eine hohe Zentralität. Die Stellung von Moskau und St. Petersburg wird von Friedmann (1995) thematisiert. Zu beachten ist, dass dieses Werk nur vier Jahre nach der Auflösung der Sowjetunion und fünf Jahre nach der deutschen Wiedervereinigung erscheint. Er vermutet darin die Etablierung von Moskau, Berlin und St. Petersburg als *world cities*. Die Evidenz dafür ist in den neueren Studien jedoch gering. Die Schwäche von Berlin könnte von der Polyzentrität Deutschlands beeinflusst sein. So gehören mit München, Frankfurt und Düsseldorf bereits drei deutsche Städte zu den zentralsten Städten Europas; so viele wie in keinem anderen europäischen Land.

Neben Europa und insbesondere Westeuropa profiliert sich der nordamerikanische Kontinent als zweite Ballungsregion globaler Städte. Tabelle 5 illustriert die Zentralitätswerte ausgewählter nordamerikanischer Städte.

⁴ Farbcodierung wird in Anhang 1 erklärt.

Tabelle 5: *Global cities und World cities in Nordamerika*

	<i>Global cities</i>				<i>World cities</i>								
Stadt	GaWC-Forschungsgruppe				Wall und v.d. Knaap				Alderson und Beckfield				Friedmann
	<i>Global connectivity</i>				OD	ID	OD	ID	OD	ID	OD	ID	
	2000	2004	2008	2010	2005				2000	2007			1995
New York	α++	α++	α++	α++	1	1	1	1	2	1	1	1	Gfa
Chicago	α	α	α	α+	14	19	20	29	13	6	11	7	Ra
Toronto	α	α	α	α		9		11	18	11	10	6	Ra
San Francisco	α	α-	β+	α					10	15	7	12	Ra
Los Angeles	α	α	α-	α					34	5		5	Mna
Philadelphia	γ	γ+	γ	α-	22				21	13	12	10	
Houston	β-	β-	β-	β+	15	14	14	10	27	22	14	13	Ra
Dallas	β-	β-	β	α-	19				15	19			
Detroit	γ+	γ	γ-	β-			21		16	21			
Boston	β	β-	β	α-		28			44	33			Ra
Atlanta	α-	β	β+	α-	29	27		27		24			
Washington	β+	β	β+	α						23			
Miami	α-	α-	β-	α-						49			Mna
Montreal	β	β	γ+	β+			25		50				Ra
Seattle	β-	γ+	γ+	β					37	50			Ra
Vancouver	β-	β-	γ+	β+									Ra

New York ist unangefochten die zentralste Stadt Nordamerikas: Friedmann kategorisiert sie als *global financial articulation city*, die zentralste Kategorie seiner Taxonomie. Sowohl die *indegrees* als auch die *outdegrees* sind auf dem höchsten Niveau und nach der GaWC-Forschungsgruppe sind New York und London die einzigen Alpha++ Städte der Welt. Hohe Zentralität und Evidenz zeigen auch Chicago und Toronto. Vergleichbar zentral sind San Francisco und Los Angeles, dafür gibt es jedoch geringere Evidenz. Nach Wall und van der Knaap rangieren diese Städte aufgrund ihrer Zentralität weder als *global cities* noch als *world cities* unter den ersten 30 weltweit. Sowohl nach der GaWC-Forschungsgruppe als auch nach Alderson und Beckfield ist ihre Zentralität hoch. Nach Friedmann zählt Los Angeles zu den *multinational articulation cities*, der zweithöchsten Kategorien seiner Taxonomie. Die Zentralität von Philadelphia, Houston, Dallas, Atlanta sowie Detroit und Boston ist geringer verglichen mit den oben erwähnten Städten. Auch die Evidenz für die Zentralität ist geringer und stammt mehrheitlich aus den Studien von Alderson und Beckfield sowie von Wall und van der Knaap. Für Washington und Montreal ergibt sich die Evidenz vor allem aus den Studien der GaWC-Forschungsgruppe. Die Zentralität von Miami ist besonders kontrovers: einerseits schätzen sowohl die GaWC-Forschungsgruppe als auch Friedmann die Zentralität als hoch ein, andererseits ist die Zentralität Miamis nach Wall und van der Knaap sowie nach Alderson und Beckfield tief. Seattle und Vancouver sind nach Friedmann *regional articulation cities*. Diese Erkenntnis findet keine Bestätigung in den neueren Studien.

Die dritte Ballungsregion an globalen Städten ist Ost- und Südostasien (Tabelle 6). Die Eruierung der zentralsten Stadt dieser Region ist sehr kontrovers. Tokyo zählt nach Friedmann mit London und New York zu den drei zentralsten *world cities* der Welt (*global financial articulation*). Diese Erkenntnis wird sowohl in den Studien von Wall und van der Knaap als auch von Alderson und Beckfield über die *world cities* bestätigt. Beide Forschungsgruppen heben insbesondere die starke Kontrollfunktion (*outdegree*) Tokyos hervor. In den Studien über die *global cities* zählt Tokyo zwar zu den zehn zentralsten Städten, jedoch nicht zur Spitze. Nach der GaWC-Forschungsgruppe profiliert sich hingegen Hong Kong als führende *global city* der Region. Nach Wall und van der Knaap sowie nach Alderson und Beckfield hat Hong Kong ein vergleichbares Prestige (*indegree*) wie Tokyo, aber eine geringe Kontrollfunktion (*outdegree*). Nach Friedmann ist Hong Kong nur eine *regional articulation city*. Singapur profiliert sich als drittes mögliches Zentrum. Nach Wall und van der Knaap hat Singapur ein sehr hohes Prestige. Nach Friedmann ist Singapur eine *multinational articulation city* und in der Stellung nur geringer als Tokyo in der Region.

Tabelle 6: Global cities und world cities in Asien

Stadt	Global cities				World cities									
	GaWC-Forschungsgruppe				Wall und v.d. Knaap				Alderson und Beckfield				Friedmann	
	Global connectivity				OD ID OD ID				OD ID OD ID					
	2000	2004	2008	2010	2005				2000		2007		1995	
Tokyo	α+	α+	α+	α+	8	6	3	6	1	4	3	4	Gfa	
Hong Kong	α+	α+	α+	α+		4		4		10		8	Ra	
Singapur	α+	α+	α+	α+	30	3		3		9		11	Mna	
Seoul	β+	α-	α	α		23		30	31	29			Ina	
Bangkok	α-	α-	α-	α-		11		18		25				
Jakarta	α-	α-	α-	α		15		15		31				
Taipei	α-	α-	α-	α-				26		37				
Peking	β+	α-	α+	α					24	36				
Kuala Lumpur	α-	α-	α	α						32				
Osaka		γ-		β-	21		28		9	39	13		Ra	
Shanghai	α-	α-	α+	α+										
Dubai	α	α	α+	α+										
Mumbai	α-	α-	α	α										
New Delhi	β	β-	β+	α-										
Manila	β+	γ+	β+	β+						42				

Hinter diesen drei Städten folgen Seoul, Bangkok, Jakarta, Taipei sowie Peking und Kuala Lumpur mit einer mittleren Zentralität und Evidenz sowohl für *global cities* als auch *world cities*. Entscheidend ist für diese Städte vor allem das Prestige (*indegree*). Osaka gilt als *world city*. Shanghai und Dubai als *global cities* nach der GaWC-Forschungsgruppe. Sie unterscheiden sich jedoch nicht in der Zentralität. Dubai, Mumbai und New Delhi sind die einzigen asiatischen globalen Städte ausserhalb der ost- und südostasiatischen

Region und gewinnen mit Peking und Shanghai nach der GaWC-Forschungsgruppe zwischen 2000 und 2010 an Zentralität. Daraus lassen sich zwei Trends für die Region ableiten: erstens, weitet sich mit Dubai, Mumbai und New Delhi die Region geographisch aus, zweitens, verstärkt sich die Zentralität von verschiedenen Städten in der Region, insbesondere in China. Manila ist zwar noch sehr peripher. Es zeigen sich jedoch erste Ansätze zur Entwicklung sowohl zu einer *global city* als auch einer *world city*.

Ozeanische, süd-, zentralamerikanische und insbesondere afrikanische Städte spielen eine untergeordnete Rolle im globalen Städtenetz. Nur wenige Städte weisen eine global relevante Zentralität auf und diese ist vom Prestige (*indegree*) gegeben. Die Kontrollfunktion (*outdegree*) ist vernachlässigbar. Tabelle 7 fasst die globalen Städte dieser drei Kontinente zusammen. Sydney und Melbourne profilieren sich sowohl als *global cities* als auch als *world cities* auf dem ozeanischen Kontinent. In Süd- und Zentralamerika besteht die Evidenz für die mittlere Zentralität von Sao Paolo, Mexico City und Buenos Aires. Geringer ist die Zentralität und ist Evidenz für Santiago, Rio de Janeiro und Karakas. Weder Friedmann noch Wall und van der Knaap noch Alderson und Beckfield erwähnen afrikanische Städte. Einzig Johannesburg und Kairo erreichen nach dem Ansatz der GaWC-Forschungsgruppe eine geringe global relevante Zentralität.

Tabelle 7: Global cities und world cities in Ozeanien, Süd-, Zentralamerika und Afrika

	Global cities					World cities							
Stadt	GaWC-Forschungsgruppe				Wall und v.d. Knaap				Alderson und Beckfield				Friedmann
	Global connectivity				OD	ID	OD	ID	OD	ID	OD	ID	
	2000	2004	2008	2010	2005				2000		2007		1995
Ozeanien													
Sydney	α	α	α+	α +		23		25		17			INA
Melbourne		α-	β+	α-				21	38	38			
Süd- und Zentralamerika													
Sao Paolo	α	α	α-	α		22		20		18			INA
Mexiko City	α	α-	α-	α		12		16		16			INA
Buenos Aires	α	α-	α	α				22		28			
Santiago	β	β+	α-	α-									
Rio de Janeiro	γ+	β-	β-	β-						46			
Karakas	β		α-	β						45			
Afrika													
Johannesburg	β+	β+	β+	α-									
Kairo	β	β	β	β+									

Neben der geographischen Verteilung der globalen Städte ist auch die Frage nach dem Kern des globalen Städtenetzes weiterhin offen. Tabelle 8 zeigt in der gewohnten Form die Zentralitätsmasse für die vier Städte, die nach dem Studium der Literatur zum Kern des globalen Städtenetzes zählen könnten. Zusätzlich zu den bisher betrachteten Studien ist auch das Werk über die *global cities* von Sassen (1991) aufgelistet. Sassen und Friedmann stimmen überein und definieren New York, London und Tokyo als Kern des

globalen Städtenetzes. Die zentrale Stellung von New York wird in allen betrachteten empirischen Studien bestätigt. London hingegen profiliert sich stärker in den *global cities*- als in den *world cities*-Ansätzen. Als *world city* weist sie zwar ein sehr hohes *indegree* auf, rangiert aber beim *outdegree* nur an vierter Stelle. Das Gegenteil gilt für Tokyo. Tokyo zeigt hohe Zentralitätswerte als *world city* und insbesondere beim *outdegree*. Paris wird weder von Sassen noch von Friedmann zu den zentralsten Städten gezählt, weist jedoch insbesondere als *world city* hohe Zentralitätswerte auf. In dieser Hinsicht gilt in Ansätzen, dass sich in Europa London und Paris die Rolle teilen: London als *global city* und Paris als *world city*.

Tabelle 8: Der Kern des globalen Städtenetzes

Stadt	Global cities						World cities							
	Sassen (1991)	GaWC-Forschungsgruppe				Wall und v.d. Knaap				Alderson und Beckfield				Friedmann
		Global connectivity				OD	ID	OD	ID	OD	ID	OD	ID	
	1991	2000	2004	2008	2010	2005				2000		2007		1995
New York	x	α++	α++	α++	α++	1	1	1	1	2	1	1	1	Gfa
London	x	α++	α++	α++	α++	3	2	4	2	4	2	4	2	Gfa
Tokyo	x	α+	α+	α+	α+	8	6	3	6	1	4	3	4	Gfa
Paris		α+	α+	α+	α+	2	5	2	5	3	3	2	3	Ina

Neben der Bedeutung der einzelnen Städte, also deren vertikaler, hierarchischer Organisation ist zudem deren horizontale Organisation von Relevanz. Insbesondere stellt sich die Frage, welche Städtesysteme, definiert als eng integrierte Gruppen von Städten, sich im globalen Städtenetz herauskristallisieren. Aus den Resultaten von Taylor et al. (2002a), Derudder et al. (2003) und Rozenblat et al. (2010) zeigt sich, dass die geographische Distanz eine entscheidende Rolle bei der Städtesystembildung spielt. Dementsprechend stimmen Städtesysteme meisten mit einer geographischen Region überein. Der Vergleich der Resultate zeigt:

- Die Städte in den USA, Europa und Ost- und Südostasien bilden drei zentrale Städtesysteme.
- Auf einem tieferen Aggregationsniveau bilden die japanischen Städte ein von den weiteren ost- und südostasiatischen Städten getrenntes System. Das europäische Städtesystem beinhaltet ein deutsches, osteuropäisches, französisches und ein britisch-irisches Städtesystem. (Derudder et al. 2003)
- Kanada zählt je nach Ansatz zu einem nordamerikanischen Städtesystem (Rozenblat et al. 2010) oder zu einem Commonwealth-Städtesystem, das neben Kanada auch Ozeanien und Südafrika umfasst (Derudder et al. 2003; Taylor et al. 2002a).
- Indische Städte einerseits und süd- und zentralamerikanische Städte andererseits bilden zwei weitere, schwach integrierte Städtesysteme. Teilweise werden die

süd- und zentralamerikanischen Städte aber mit den westeuropäischen Städten in ein übergreifendes System integriert. (Derudder et al. 2003)

— Afrikanische Städte werden, wenn überhaupt, ins europäischen System integriert (Rozenblat et al. 2010).

Ein Punkt, der wie erwähnt bisher noch wenig Beachtung in empirischen Studien gefunden hat, ist die Berücksichtigung von nicht-ökonomischen Kriterien für die Definition der globalen Städte. Tabelle 9 stellt eine Taxonomie der globalen Städte aufgrund ihrer wirtschaftlichen, politischen, gesellschaftlichen und kulturellen Interaktionen nach Taylor (2005a) dar. Gemäss diesem Ansatz sind London und New York die funktionsübergreifenden und führenden globalen Städte. Den funktionsübergreifenden globalen Städten gehören Los Angeles, San Francisco, Boston, Chicago und Toronto in Nordamerika sowie Paris, Amsterdam, Madrid, Mailand und Moskau in Europa an. Bei der wirtschaftlichen Funktion profilieren sich die drei ost- und südasiatischen Städte, Hong Kong, Singapur und Tokyo sowie auf tieferer hierarchischer Ebene auch Miami, Frankfurt, Zürich, Osaka und Sydney. Afrikanische, süd- und zentralamerikanische Städte kommen nicht vor.

In sozio-politischer Hinsicht profilieren sich Brüssel, Genf und Washington. Hinzu kommen als politische Zentren Bangkok, Peking und Wien. Zu den gesellschaftlichen Zentren gehören Manila, Nairobi und Ottawa. In kultureller Perspektive kommen Berlin, Kopenhagen, Melbourne, München, Oslo, Rom, und Stockholm hervor. Zudem weisen Los Angeles, Paris und San Francisco einen kulturellen Schwerpunkt auf, obwohl sie funktionsübergreifende Städte sind. Bei der Betrachtung der nicht-wirtschaftlichen Funktionen werden zudem noch Abidjan, Addis Abeba, Atlanta, Basel, Barcelona, Kairo, Denver, Harare, Lyon, Manila, Mexiko City, Mumbai, New Delhi und Shanghai erwähnt.

Tabelle 9: Taxonomie der Städte unter Berücksichtigung der wirtschaftlichen, politischen, gesellschaftlichen und kulturellen Funktionen (Anpassung von Taylor 2005a)

	Wirtschaft	Politik	Gesellschaft	Kultur
Führende funktionsübergreifende globale Städte (<i>functionally comprehensive global cities</i>)	London, New York Los Angeles, Paris, San Francisco (schwacher kultureller Schwerpunkt) Amsterdam, Boston, Chicago, Madrid, Mailand, Moskau, Toronto			
Spezialisierte führende globale Städte (<i>global niche cities</i>)	Hong Kong, Singapur, Tokyo	Brüssel, Genf, Washington		
Führende untergeordnete Weltstädte (<i>subnet articulator cities</i>)		Bangkok, Peking, Wien	Manila, Nairobi, Ottawa	Berlin, Kopenhagen, Melbourne, München, Oslo, Rom, Stockholm
Weltweit führende Städte (<i>worldwide leading cities</i>)	Frankfurt, Miami, Osaka, Sydney, Zürich	Abidjan, Addis Abeba, Atlanta, Basel, Barcelona, Kairo, Denver, Harare, Lyon, Manila, Mexiko City, Mumbai, New Delhi, Shanghai		

Daraus lässt sich schliessen, dass sich unter Berücksichtigung von politischen, gesellschaftlichen und kulturellen Funktionen die Geographie des globalen Städtensetzes nicht massgeblich vom wirtschaftlichen Städtensetz unterscheidet. Drei wichtige Punkte sind trotzdem erwähnenswert:

- Durch den Einbezug von nicht-ökonomischen Faktoren wird die Stellung der afrikanischen Städte zwar gestärkt, dennoch haben sie einen geringen Einfluss verglichen mit den europäischen, nordamerikanischen, sowie ost- und südostasiatischen Städten im globalen Netzwerk.
- In sozio-politischer Hinsicht werden insbesondere Hauptstädte oder Sitze von internationalen Organisationen und Botschaften gestärkt.
- Bei der kulturellen Funktion liegt der geographische Schwerpunkt bei den europäischen Städten.

Während über das wirtschaftliche globale Städtensetz eine Vielzahl von Studien existieren, wovon hier nur eine Auswahl präsentiert wird, beschränkt sich die multifunktionale Untersuchung des multifunktionalen globalen Städtensetzes auf diese Studie von Taylor (2005a). Diese Arbeit greift dieses Manko auf und präsentiert neben einem methodischen Rahmen (Kapitel 3) auch neue Erkenntnisse über das multifunktionale globale Städtensetz (Kapitel 7.3).

2.4 Fazit

Städte sind fokale Knoten der Informationsgesellschaft und ersetzen nach Taylor (2000) den Nationalstaat als Kern der modernen Gesellschaft. Der Globalisierungsprozess und die technologische Entwicklung führen zur Entwicklung eines globalen Interaktionsnetzwerks zwischen den Städten. Das globale Städtensetz entsteht durch die dynamische Wechselwirkung zwischen der Stadt und den urbanen Akteuren. Obwohl sowohl auf theoretischer als auch auf empirischer Ebene der Fokus auf den wirtschaftlichen Beziehungen zwischen den Städten liegt, postulieren mehrere Ansätze die multifunktionale (politische, ökonomische, kulturelle und gesellschaftliche) Einbettung der Stadt ins globale Netzwerk. Dieser multifunktionale Ansatz liegt auch der vorliegenden Arbeit zugrunde.

Auf methodischer Ebene fasst das Zitat von Short (1996) „*The dirty little secret of world cities research*“ die Problematik bestens zusammen. Die etablierten Datenquellen, im Speziellen die angenommenen Interaktionen zwischen den APS-Niederlassungen, die Besitzverhältnisse innerhalb der transnationalen Unternehmen, oder die Flugpassagierbewegungen bilden nur die globale und relationale Natur der betrachteten urbanen Prozesse ab. Die Multifunktionalität bleibt damit weitgehend unberücksichtigt.

Die methodischen Ansätze stützen sich mehrheitlich auf netzwerkanalytische Methoden. Nur bei wenigen Untersuchungen des globalen Netzwerkes wird das Potential der visuell-analytischen Analysen als Ergänzung zur Netzwerkanalyse erkannt, jedoch nicht

ausgeschöpft. Ein kohärenter und theoretisch fundierter methodischer Rahmen, der die netzwerk- und die visuell-analytischen Ansätze für die Untersuchung des globalen Städtetetzes integriert, ist bislang nicht vorhanden und bildet die zweite methodische Herausforderung dieser Arbeit.

Wegen den erwähnten offenen Fragen im methodischen Bereich liegen nur wenige Resultate über die soziale, politische und kulturelle Einbettung der Städte im globalen Netzwerk vor. Das ökonomische Städtetetz ist hingegen vertieft untersucht worden und dient als Referenz für die Validierung der inhaltlichen Resultate dieser Arbeit. Aus wirtschaftlicher Perspektive profilieren sich Nordamerika, Europa sowie Ost- und Südostasien als Ballungsregionen. New York, London, Paris und Tokyo treten als Kern des globalen Städtetetzes hervor.

3 Methodentheoretischer Rahmen

Das folgende Kapitel entwickelt den methodentheoretischen Rahmen für die Exploration und Analyse von Netzwerkdaten, welcher sich auf das *spatialization framework* von Fabrikant und Skupin (2005; 2007) stützt. Das *spatialization framework* kombiniert Raumanalyse mit Visualisierung zu einem (geo-)visuell-analytischen Ansatz und ermöglicht zudem die Integration von Methoden und Ansätzen aus verschiedenen Disziplinen. Der Fokus des ersten Teils des Kapitels liegt auf konzeptionell-theoretischer Ebene. Dabei werden ausgewählte Aspekte der Forschung zu sozialen Netzwerkanalyse, zur Informationsvisualisierung und Kartographie durchleuchtet und zu einem kohärenten Rahmen verbunden. Im zweiten Teil werden die theoretischen Erkenntnisse in einen konkreten Arbeitsablauf umgesetzt. Darauf stützt sich das methodische Vorgehen der Arbeit.

3.1 *Spatialization framework*

Das populäre Sprichwort *Ein Bild sagt mehr als tausend Wörter* unterstreicht die Macht der graphischen Darstellungen bei der Informationsvermittlung. Mit der graphischen Darstellung der Daten werden die kognitiven Fähigkeiten der Lesenden unterstützt und sogar erweitert (Tversky 2000), und letztlich die Exploration und die Analyse der Daten vereinfacht (Fabrikant und Skupin 2005).

Verschiedene Disziplinen beschäftigen sich mit der Visualisierung von Daten. Die Kartographie (Bertin 1974; MacEachren 1995) befasst sich mit der salienten Darstellung von räumlichen Daten in Form einer Karte. Die Informationsvisualisierung (*information visualization*) (Card et al. 1999; Börner et al. 2003; Chen 2003, 2004) ist bei der Visualisie-

rung von nicht-räumlichen Daten die führende Disziplin. Das *spatialization Framework* beschäftigt sich mit der Visualisierung von nicht-räumlichen Daten und integriert dabei Ansätze und Methoden aus beiden Disziplinen.

Die Informationsvisualisierung, als wissenschaftlicher Prozess, wird als „*use of computer-supported, interactive, visual representations of abstract data to amplify cognition*“ (Card et al. 1999) definiert. Dabei lassen sich drei wesentliche, definitorische Elemente identifizieren: die abstrakte, nicht-räumliche Natur der Ausgangsdaten, deren graphische Darstellung, sowie die Erweiterung des Wahrnehmungsvermögens des Nutzenden, um die Exploration und Analyse der Daten zu erleichtern. Das *spatialization framework* wird nach Skupin und Fabrikant (2007) als „*systematic transformation of high-dimensional data sets into lower-dimensional, spatial representations for facilitating data exploration and knowledge construction*“ definiert. Das zu erreichende Ziel („*facilitating data exploration and knowledge construction*“) ist dasselbe wie bei der Informationsvisualisierung, unterschiedlich ist jedoch der Prozess, der dazu führt. Die *spatialization* wird als systematischer (Skupin und Fabrikant 2007) und formalisierter (Fabrikant und Skupin 2005) Prozess definiert. Systematisch bezieht sich auf die klar definierten Arbeitsschritte, welche zur Visualisierung führen. Formalisiert bedeutet, dass der Visualisierung formale Regeln zugrunde liegen.

Das systematische Vorgehen und die formellen Regeln stammen mehrheitlich aus etablierten Theorien und Methoden der Kartographie, wie der Lehre der graphischen Variablen (Bertin 1974; MacEachren 1995) und dem Generalisierungsprozess (McMaster 1989; Buttenfield und McMaster 1991). In der Informationsvisualisierung wird der Einfluss der kartographischen Tradition anerkannt (Chen 2003), jedoch wird diese Tradition selten systematisch einbezogen und umgesetzt (Skupin 2000; Skupin und Fabrikant 2003).

Die Erweiterung des Wahrnehmungsvermögens wird durch den Einsatz von Metaphern, insbesondere räumlichen Metaphern (Chen 2004) unterstützt. Skupin (2000) präzisiert jedoch, dass es nicht genügt, dass Informationsvisualisierungen wie Karten aussehen, sondern der kognitive Mehrwert wird nur durch das Anwenden von räumlichen Gesetzen erreicht. Couclelis (1998) doppelt nach und unterstreicht die Zentralität der räumlichen Metaphern bei der Konzipierung der Informationsvisualisierung. Räumliche Metaphern schöpfen ihre Bedeutung aus dem physisch existierenden geographischen und dem wahrgenommenen geographischen Raum. Diese werden in unterschiedlichen ontologischen Arbeiten (Lakoff und Johnson 1980; Golledge 1995; Couclelis 1998; Fabrikant 2000) definiert und dienen als Grundlage bei der Konzipierung der *spatialization*.

Der *spatialization*-Prozess kann analog zum kartographischen Generalisierungsprozess in die semantische und die geometrische Generalisierung unterteilt werden. Die semantische Generalisierung agiert auf der theoretisch-konzeptionellen Ebene. Sie umfasst die Selektion der darzustellenden Daten (Robinson et al. 1978) und die metaphorische Kartierung. Mit der Selektion werden die Hauptmerkmale des zu analysierenden Prozesses oder Ereignisses identifiziert und die entsprechenden Daten ausgewählt. Die Selektion

erfolgt in Abhängigkeit von Massstab und Ziel. Die Objekte, deren räumliche, zeitliche und inhaltliche Eigenschaften, sowie deren Granularität werden Zielbereich (*target domain*) genannt. Die metaphorische Kartierung verbindet den Zielbereich mit dem wahrgenommenen geographischen Raum (Quellbereich, *source domain*). Räumliche Metaphern, welche den semantischen Zielbereich möglichst präzise abbilden, werden formalisiert. Zum Beispiel kann die semantische Distanz zwischen zwei Objekten im Informationsraum metaphorisch in Bezug zur Distanz zwischen zwei Ortschaften im geographischen Raum gesetzt werden, oder die Grenzen zwischen Datenclustern beziehen sich auf die Grenzen zwischen Regionen. (Fabrikant und Buttenfield 2001; Fabrikant und Skupin 2005)

Mit der geometrischen Generalisierung wird eine graphische Darstellung generiert, welche die Daten den Metaphern entsprechend darstellt (Fabrikant und Skupin 2005). Im Unterschied zur kartographisch-geometrischen Generalisierung gehen bei der *spatialization* den graphischen Operationen mathematische Berechnungen voraus, mit denen die Geometrie und die Attribute generiert werden. Dieser Unterschied lässt sich sehr gut an einem Beispiel zeigen. In topographischen Karten ist die Lage einer Stadt durch die gemessenen geographischen Koordinaten gegeben und kann dementsprechend platziert werden. Die Lage eines Textes im Informationsraum hängt von dessen inhaltlicher Ähnlichkeit zu anderen Texten im Korpus ab und ist per se nicht gegeben. Die Koordinaten müssen mathematisch, z.B. mit Dimensionsreduktionsverfahren berechnet werden und können erst danach dargestellt werden. Solche mathematischen Transformationen sind für die Generierung jedes Attributs notwendig. Wegen des grossen Stellenwertes der mathematischen Transformationen in dieser Arbeit wird der betreffende Arbeitsschritt als eigenständiger Prozessschritt behandelt.

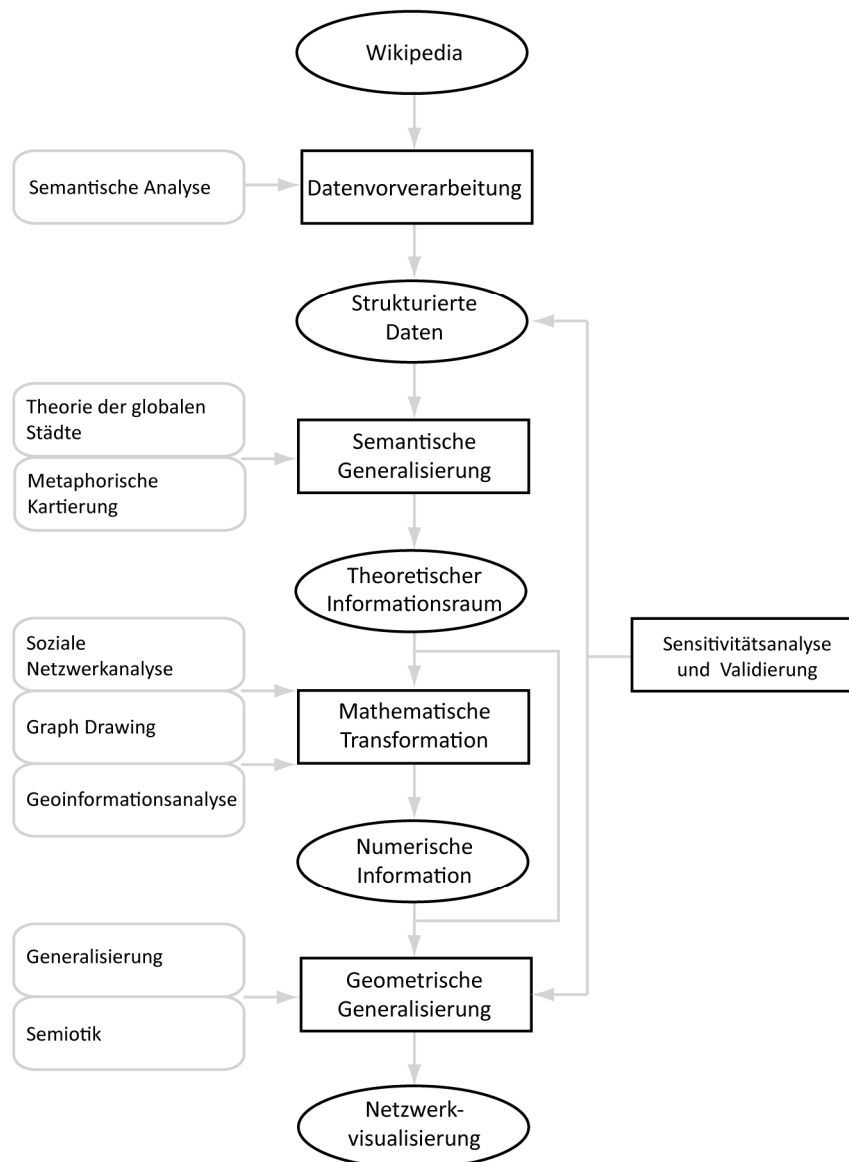
Der Kern der geometrischen Generalisierung ist die saliente graphische Umsetzung der numerischen Information, sowie das Gleichgewicht zwischen Informationsdichte und Lesbarkeit grafisch umzusetzen. Dabei werden Attribute mit graphischen Variablen dargestellt und der graphische Output wird anhand von Operationen der kartographisch-geometrischen Generalisierung dem Massstab angepasst. Von den vier Operatoren nach Robinson et al. (1978), Selektion (*selection*), Klassierung (*classification*), Vereinfachung (*simplification*), und Symbolisierung (*symbolisation*), gehören die letzten zwei zur geometrischen Generalisierung. Die Selektion ist, wie oben erwähnt, Teil der semantischen Generalisierung und die Klassierung der mathematischen Transformation.

In Skupin und Fabrikant (2003; 2007) wird ein dritter, weniger zentraler, jedoch sehr wichtiger Prozessschritt erwähnt: die Datenvorverarbeitung, welche insbesondere bei unstrukturierten Daten (z.B. Texten) unabdingbar ist. Mit diesem Schritt, welcher ganz zu Beginn des *spatialization*-Prozesses steht, werden die Daten reorganisiert und teilweise bereinigt.

In den folgenden Unterkapiteln werden die Prozessschritte mit Bezug auf den aktuellen Stand der Forschung wie folgt auf die Fragestellung spezifiziert (Abbildung 10):

- Datenvorverarbeitung: Die verwendete Datenbasis Wikipedia besteht aus nutzergenerierten, halbstrukturierten Textdaten. Diese müssen zuerst durch die Datenvorverarbeitung strukturiert werden.
- Semantische Generalisierung: Die Theorie der *global cities* ist ein netzwerktheoretischer Ansatz. Diese Perspektive muss auch bei der Konzeption des Informationsraumes mittels metaphorischer Kartierung im Vordergrund stehen.
- Mathematische Transformationen: Die Formalisierung des theoretischen Informationsraums durch mathematische Information stützt sich auf die Integration von sozialer Netzwerkanalyse, *graph drawing* und Geoinformationsanalyse.
- Geometrische Generalisierung: Die Grösse der Datensätze und deren Komplexität stellen besondere Anforderungen an die Visualisierung. Die systematische Anwendung der Zeichenlehre und der kartographischen geometrischen Generalisierung erhöht die Salienz der Netzwerkvisualisierung.
- Resultate: Sowohl die strukturierten Daten, die aus der Datenvorverarbeitung stammen, als auch die numerischen Informationen, welche mathematisch hergeleitet werden, sind durch eine gewisse Unsicherheit gekennzeichnet. Verzerrungen können sowohl intrinsisch in den Daten vorhanden sein oder durch die gewählten Methoden und Parameter generiert werden. Sensitivitätsanalysen und Validierungen der Resultate sind somit von besonderer Bedeutung.

Abbildung 10: Übersicht des methodentheoretischen Rahmens mit den verschiedenen Prozessschritten und den Resultaten, die sich daraus ergeben



3.1.1 Semantische Generalisierung

Das Ziel jeder *spatialization* ist, den Zugang zu den Daten zu erleichtern und dadurch die Datenexploration und -analyse zu fördern, um neue Erkenntnisse zu gewinnen. Wenn die graphische Darstellung auf die menschlichen Visualisierungskapazitäten abgestimmt ist, sind die wichtigsten Informationen einfach und intuitiv erkennbar (Fabrikant und Skupin 2005). Diese kognitiv inspirierte Konzeption der Visualisierung findet in der semantischen Generalisierung mittels metaphorischer Kartierung (*metaphorical mapping*) statt. Jede Metapher hat einen Quell-, einen Zielbereich und eine Quell-Ziel-Kartierung. Durch die metaphorische Kartierung wird der abstrakte Inhalt des Zielbereiches dank einer Metapher, welche ihre Aussagekraft aus dem konkreten Quellbereich schöpft, wiedergegeben. Die Stärke der Metapher hängt von der Erfahrung ab, die die Nutzenden mit dem Quellbereich haben. Nur wenn der Quellbereich konkret und bekannt ist, entfaltet die Metapher ihre Wirkung und der Zielbereich wird verständlicher (Johnson

1987). Der Desktop und die digitalen Ordner in einem Computer sind Beispiele zweier Metaphern. Räumliche (*spatialization*) Metaphern, welche sich auf die physikalische und kulturelle Erfahrung stützen, sind zudem besonders aussagekräftig (Lakoff und Johnson 1980). Johnson (1987) unterstreicht den Nutzen der Metaphern auch im wissenschaftlichen Denken. Fabrikant (2000) und Couclelis (1998) definieren und belegen deren Anwendung auch in der Informationsvisualisierung. Im Kontext der Informationsvisualisierung besteht der abstrakte Zielbereich aus den Daten (semantische Umwelt), welche in graphischer Form dargestellt werden. Die geographische Umwelt in ihren physischen, wahrgenommenen und erfahrenen Ausprägungen bildet den konkreten Quellbereich. Die Kartierung erfolgt mittels räumlicher Metaphern, dabei sind semantische Primitive essentielle Elemente:

- Die semantischen Primitiven sind prinzipielle Metaphern, welche zu komplexeren und bedeutungsreicheren Metaphern kombiniert werden.
- Die Daten (semantische Umwelt) werden mittels semantischer Primitiven im Informationsraum (Zielbereich) dargestellt.
- Semantische Primitive bergen die essentiellen Eigenschaften der geographischen Umwelt (Quellbereich).

Fabrikant und Skupin (2005) definieren vier semantische Primitive: Lokus, Trajektorie, Grenze und Aggregat (Tabelle 10). Diese stimmen sehr stark mit den netzwerkanalytischen Primitiven überein. Diese sind nach Wasserman und Faust (2008): Knoten, Kanten, Subgruppe oder Partition und Netzwerk (Tabelle 10).

Tabelle 10: Primitiven des Zielbereichs aus netzwerkanalytischer Perspektive und aus Sicht der *spatialization*

Netzwerkanalytische Primitiven	Semantische Primitiven
Knoten	Lokus
Kante	Trajektorie
Subgruppe / Partition	Grenze Aggregat
Netzwerk	Netzwerk

Die Bedeutung der semantischen Primitiven wird anhand der folgenden geoinformationswissenschaftlichen, linguistischen, und psychologischen Studien diskutiert (für eine Übersicht siehe Tabelle 11):

- “*The image of the city*” von Kevin Lynch (1962). Es handelt sich um eine bekannte und einfach auf den Informationsraum übertragbare Definition der geographischen Umwelt. Lynch (1962) analysiert Städte aus der Navigationsperspektive und identifiziert dabei fünf Elemente, die wiederholt in mentalen Karten vorkommen und zu denen alle Bauten und Gegenstände einer Stadt zugewiesen werden können. Fabrikant und Skupin (2005) assoziieren vier von diesen Ele-

menten (Brennpunkte bleiben unberücksichtigt) mit den semantischen Primitiven. Die Navigationsperspektive ist nur eine mögliche Betrachtung des geographischen Raumes (für weitere Perspektiven siehe Couclelis 1992; Fabrikant und Skupin 2005).

- “*Primitives of spatial knowledge*” von Golledge (1995). Golledge (1995) präsentiert eine minimale Menge von grundsätzlichen räumlichen Konzepten, welche auch ohne vertiefte Ausbildung verstanden werden.
- Die kognitiv-linguistische Definition der Image-Schemas von Johnson und Lakoff (Johnson 1987; Lakoff 1987). Image-Schemas werden von Johnson (1987) als “*embodied patterns of meaningfully organized experience*” definiert und sind somit die grundlegenden Strukturen unserer Erfahrung. Deren Einbeziehung in die metaphorische Kartierung bereichert die Metaphern um den Bereich der Erfahrung und erhöht ihre Aussagekraft beträchtlich (Couclelis 1998; Fabrikant 2001a).
- Gestaltgesetze der Berliner Schule der Gestaltpsychologie, insbesondere von Max Wertheimer, Wolfgang Köhler und Kurt Koffka (z.B. Koffka 1935). Diese Gesetze sind ein fundamentaler Beitrag zur Wahrnehmung von Mustern in Visualisierungen (MacEachren 1995; Ware 2004). Diese Prinzipien werden bisher selten systematisch auf die Visualisierung von Netzwerken angewendet (Ware 2004).

Tabelle 11: Metaphorische Kartierung. Assoziierung des Zielbereichs (netzwerkanalytische und semantische Primitiven) mit dem Quellbereich definiert anhand der geographischen Primitiven in Navigationsperspektive nach Lynch (1962), den Image-Schemas nach Johnson und Lakoff (Lakoff und Johnson 1980; Lakoff 1987; Johnson 1987) und den Primitiven des räumlichen Wissens nach Golledge (1995)

Semantische Primitiven	Lokus	Trajektorie	Aggregat / Grenze	Netzwerk
Kognitive Konzepte				
Netzwerkanalytische Primitiven	Knoten	Kante	Subgruppe / Partition	Netzwerk
Geographische Primitiven (Navigationsperspektive)	Brennpunkt	Weg	Bereich Grenzlinie	Umwelt (Stadt)
Gestaltgesetze	Nähe	Verbundenheit, Kontinuität	Gemeinsame Region	
Image-Schemas und Primitiven des räumlichen Wissens				
Grundlegende Konzepte	Identität Lage Grösse	Verbindung Link	Container	Hierarchie Zentrum- Peripherie Korrelation Teil-Ganzes
Abgeleitete Konzepte	Distanz Nah-Fern	Distanz Pfad (Richtung, Abfolge)	Dichte Verteilung Form / Muster	

Lokus

Knoten und Loci sind beide punktuelle Elemente. Das geographische Äquivalent sind Brennpunkte, Punkte intensiver Aktivität. Typisches Beispiel sind Plätze. Sowie ein Platz durch einen Namen und eventuell seine Grösse gekennzeichnet ist, so besitzen Knoten im Informationsraum auch eine Grösse und eine Bezeichnung, meistens in der Form einer Beschriftung. Objekte werden aus geographischer Perspektive massgeblich durch die Lage gekennzeichnet, analog dazu ist auch die Lage die prägendste Eigenschaft eines Knotens im Informationsraum. Die Lage im Informationsraum ist relativ und durch die Beziehung von jedem einzelnen Knoten zu den anderen gegeben. Sobald zwei Objekte eine Lage besitzen, lässt sich deren Distanz ableiten (Golledge 1995).

Die Bedeutung von Lage und Distanz kann vollumfänglich nur durch den Einbezug des ersten Gesetzes der Geographie verstanden werden: „*Everything is related to everything else, but near things are more related than distant things*“ (Tobler 1970). Mathematisch spricht man von räumlicher Autokorrelation. Diese Eigenschaft liegt der Distanz-Ähnlichkeitsmetapher (Montello et al. 2003) zugrunde. Demzufolge sind naheliegende Knoten im Informationsraum einander ähnlich. Die Distanz-Verbundenheitsmetapher (Salvini et al. 2011) ist eine Anpassung davon für Informationsräume, in denen die Lage der Knoten von der gegenseitigen Verbundenheit und nicht von der Ähnlichkeit der Elemente gegeben ist.

Montello et al. (2003) bestätigen empirisch die kognitive Gültigkeit der metaphorischen Eigenschaft der Loci im Informationsraum und sprechen in dieser Hinsicht vom ersten Gesetz der kognitiven Geographie, welches besagt, dass „*people believe closer things are more similar*“. Dies stimmt mit dem Nähe-Distanz-Image-Schema und dem Gestaltgesetz der Nähe überein. Dem ersten zufolge werden nahe Objekte als relevanter (Fabrikant und Buttenfield 2001), dem zweiten zufolge als zusammengehörend wahrgenommen (Ware 2004).

Trajektorie

Kanten und Trajektorien sind lineare Elemente. Die Quelle für diese semantische Primitive sind Wege. Die grundlegende Bedeutung dieser Konzepte ist das Verbinden: Kanten verbinden Knoten, Trajektorien verbinden Loci und Wege verbinden Orte.

Trajektorien prägen die Wahrnehmung des Informationsraums, sie betonen und ergänzen die Distanz-Ähnlichkeitsmetapher bzw. die Distanz-Verbundenheitsmetapher. Dies kann aus unterschiedlichen Perspektiven erklärt werden. Verbindungen sind prägende Elemente des geographischen Wissens (Golledge 1995) und somit Teil unserer Erfahrung. Gemäss dem Link-Image-Schema führen Verbindungen auf Gemeinsamkeiten zurück (Johnson 1987). Wahrnehmungspsychologisch bestätigen Palmer und Rock (1994), dass Verbindungen ein prägendes organisatorisches Element sind, welches die direkte Distanz sogar überspielt (Ware 2004).

Aus den Trajektorien lässt sich auch das Konzept des Pfads ableiten. Pfade werden als von einem Start- zu einem Zielpunkt führende Wege empfunden (Lakoff 1987). Ob-

wohl Pfade nicht inhärent gerichtet sind, werden sie als solche wahrgenommen (Johnson 1987). Obwohl die Konzepte der Richtung und Abfolge nicht intrinsisch in ungerichteten Netzwerken vorhanden sind, werden sie in deren Visualisierung wahrgenommen. Anders als im geographischen Raum, in dem Richtung und Abfolge unabhängig von den Wegen existieren, manifestiert sich diese nur als empfundene Eigenschaft der Pfade.

In Visualisierungen mit Trajektorien wird die direkte Distanz durch zwei zusätzliche Distanzmetriken ergänzt: die Distanz entlang eines Pfades und die topologische Distanz. Die drei Konzepte lassen sich gut mit einem geographischen Beispiel erklären. Die direkte Distanz ist die Luftlinie zwischen einem Ort A und einem Ort B. Die Distanz entlang eines Pfades ist die Distanz entlang einer Strasse, welche A und B verbindet. Die topologische Distanz entspricht der Anzahl Orte, durch die man entlang der Strasse zwischen A und B fährt. (Fabrikant et al. 2004)

Die starke Aussagekraft der Verbindungen, welche bereits von Palmer und Rock (1994) postuliert wird, wird von Fabrikant et al. (2004) in Bezug auf die Einschätzung der Ähnlichkeit in Visualisierungen bestätigt. Nach dem ersten Gesetz der kognitiven Geographie führt man die Distanz auf die Ähnlichkeit zurück. In Netzwerken ist die Distanz entlang der Kanten die bestimmende Distanzmetrik. Die direkte Distanz und die topologische Distanz (Anzahl Knoten) zwischen zwei betrachteten Knoten wirken weniger stark auf die Beurteilung der Ähnlichkeit (Fabrikant et al. 2004).

Für die Visualisierung von Pfaden ist das Gesetz der Kontinuität von Bedeutung. Demzufolge werden kontinuierlich verlaufende Pfade eher als solche wahrgenommen (Ware 2004). Kantenknicken, Kantenkreuzungen und akute Ecken zwischen den Kanten verringern die Wahrnehmung der Pfade (Bennett et al. 2007).

Aggregat und Grenze

Netzwerke werden aufgrund endogener (Subgruppen) und exogener Eigenschaften (Partitionen) in Gruppen unterteilt. Das semantische Äquivalent sind Aggregate. Aggregate sind meistens flächenhafte Objekte, welche das Resultat von Klassifikationen sind. Aggregate entsprechen Bereichen, z.B. Bezirken oder Quartieren aus geographischer Perspektive. Sowohl Bereiche als auch Aggregate und Gruppen sind unter Betrachtung ausgewählter Eigenschaften homogen und zugleich verschieden von ihrer Umgebung (Fabrikant und Skupin 2005).

Die Eigenschaften der Aggregate sind am besten im Container-Image-Schema enthalten. Container implizieren ein Innen, ein Aussen und eine dazwischen verlaufende, trennende Grenze (Lakoff 1987). Grenzen sind eine weitere semantische und geographische Primitive, welche auch unabhängig vom Aggregat und Bereich existiert. Im Sinne des Container-Image-Schemas und weil in der Netzwerkanalyse das Konzept der Grenze abwesend ist, wird die Grenze gekoppelt mit dem Aggregat diskutiert. Nach Golledge (1995) sind Grenzen die Voraussetzung für die Analyse von Dichten und Verteilungen. Da die Dichte auch ein wichtiges Mass für die Analyse der Subgruppen und

Partitionen ist, sind Grenzen auch in Netzwerken vorhanden. Mit der Visualisierung dieser Grenzen durch Linien wird deren Vorhandensein betont.

Aggregate können sowohl als eine Sammlung von Objekten als auch als eine homogene Zone verstanden werden. Im ersten Fall kann man sich aus geographischer Perspektive einen Wald und im zweiten Fall einen See vorstellen (Fabrikant und Skupin 2005). Die Wahrnehmung einer Gruppe aus verschiedenen einzelnen Objekten, wird durch die Geseztgesetze beeinflusst. Naheliegende (Gesetz der Nähe), verbundene (Gesetz der Verbundenheit) Objekte werden als Teil desselben Aggregats wahrgenommen. Das Gesetz, welches am Besten die Semantik des Aggregats wiedergibt, ist jedoch das Gesetz der gemeinsamen Region (Palmer 1992). Das Gesetz hebt hervor, dass Elemente, die sich im gleichen Perimeter befinden, eher zueinander gruppiert werden (MacEachren 1995). Durch das Abgrenzen von Regionen wird die Unterscheidung zwischen innen und aussen besonders betont. Dieses Gesetz und dessen Gültigkeit bei der *spatialization* wird von Fabrikant et al. (2006) empirisch bestätigt.

Das Teil-Ganzes-Image-Schema hebt einen wichtigen Aspekt der Aggregate hervor. Aggregate sind Teil eines Ganzen, im Spezifischen des Netzwerks. Zugleich sind Aggregate ein Ganzes, welches die einzelnen dazugehörenden Objekte umfasst (Couclelis 1998). Diese Doppeldeutigkeit wird insbesondere bei Konzepten wie Dichte, Verteilung, Muster und Form (Golledge 1995) ersichtlich. Aufgrund unterschiedlicher Dichte (z.B. sehr hoher Dichte), Verteilung (z.B. Cluster), Muster und Form werden Aggregate im Informationsraum, Gruppen in Netzwerken und Bereiche im geographischen Umfeld zu Teilen eines Ganzen. Gleichzeitig setzen diese Eigenschaften auch voraus, dass die Aggregate bzw. die Gruppen ein Ganzes und die Loci bzw. die Knoten ein Teil davon sind. Enthalten und enthalten sein sind intrinsische Eigenschaften der Aggregate.

Netzwerk

Das Netzwerk ist per se keine Primitive, sondern die Gesamtheit, welche durch das Zusammenwirken aller Primitiven entsteht. Somit entspricht es keinem der semantischen Primitiven und keinem der fünf Elemente nach Lynch (1962). Couclelis (1992) ergänzt jedoch die Raumelemente aus verhaltensgeographischer Perspektive mit dem Konzept „Umwelt“, ein allumfassendes Element, analog dem Netzwerk, welches alle Wege, Grenzlinien, Bereiche und Brennpunkte umfasst. Aus Lynch's (1962) Perspektive wäre die Umwelt die Stadt.

Konzepte des räumlichen Wissens von höherer Ordnung sind in jedem Netzwerk intrinsisch vorhanden: nach Golledge (1995) entstehen Netzwerke durch das Zusammenwirken von Lage und Verbundenheit und Hierarchie durch das Vereinen von Lage, Grösse und Verbundenheit. Das Zentrum-Peripherie-Image-Schema bezieht sich auf die hierarchische Organisation der Wahrnehmung; diese stimmt mit den Grundsätzen der räumlichen Organisation überein: das Zentrum ist wichtig, die Peripherie weniger (Johnson 1987).

Neben der räumlichen Autokorrelation (siehe Lokus) ist Massstabsabhängigkeit eine weitere fundamentale Eigenschaft des geographischen Raumes (National research council 1997). Geographische Prozesse spielen sich auf sehr unterschiedlichen Massstabsebenen ab; von lokal bis global. Je nach Massstab sind andere Elemente oder andere Eigenschaften des Raumes von Relevanz. Bei der kartographischen Darstellung von geographischen Prozessen ist die Generalisierung die Antwort auf die ändernden Massstäbe. Die Massstabsabhängigkeit ist eng mit dem Vertikalität-Image-Schema verbunden. Aufgrund unserer Erfahrung ist mehr mit höher und weniger mit tiefer gekoppelt, dies sowohl in Bezug auf qualitative als auch auf quantitative Aspekte (Johnson 1987). Die Übertragbarkeit auf abstrakte Sachverhalte wird selbst von Johnson (1987) postuliert. Auf den Informationsraum übertragen wirkt sich der Massstab auf den Detaillierungsgrad aus, auf die Granularität der Darstellung, sowohl in geometrischer (*level of detail*) als auch in semantischer Hinsicht, also sowohl auf was als auch auf wie es dargestellt wird (Fabrikant 2001b; Fabrikant 2004).

3.1.2 Mathematische Transformation

Die mathematische Transformation umfasst sowohl semantische als auch geometrische Operationen. Geometrische Operationen befassen sich mit dem Kern jeder *spatialization*, der Transformation der mehrdimensionalen Inputmatrix in eine zweidimensionale räumliche Darstellung (Skupin und Fabrikant 2007). Der Forschungsbereich des *graph drawings* hat dafür sehr unterschiedliche Methoden entwickelt.

Obwohl der Inhalt der Netzwerke sehr heterogen ist und sowohl Schulkinder und Freundschaften als auch Flughäfen und Passagiere umfasst, können sie auf eine Menge von Knoten und Kanten bzw. Akteure und Beziehungen zurückgeführt werden und in wenige Typen unterschieden werden. Netzwerke können aufgrund ihrer Knoten bzw. Akteure in unimodale und bimodale Netzwerke unterschieden werden (Wasserman und Faust 2008). In unimodalen Netzwerken gehören alle Knoten zur selben Art. Typisches Beispiel sind Beziehungen zwischen Mitarbeitenden einer Firma. In bimodalen Netzwerken stehen zwei Arten von Knoten in Bezug zueinander, z.B. DorfbewohnerInnen, die unterschiedliche Vereine besuchen. Netzwerke lassen sich auch durch deren Kanten charakterisieren. Gerichtete Netzwerke werden von ungerichteten und gewichtete von ungewichteten unterschieden (Wasserman und Faust 2008).

Attribute und soziale Netzwerkanalyse

Es existiert eine breite und heterogene Palette an Massen und Algorithmen, welche auf dem Ansatz der Netzwerkanalyse beruhen (für eine Übersicht siehe Brass 1995). Dabei stehen traditionelle Masse, welche insbesondere für kleinere überschaubare, meistens interpersonelle Netzwerke entwickelt wurden, neueren Massen, welche sich insbesondere für sehr grosse und komplexe Netzwerke (z.B. das *World Wide Web*) eignen, gegenüber. Diese werden in Anlehnung an Wasserman und Faust (2008) sowie Monge und Contractor (2003) vier Analyseebenen zugewiesen: einer individuellen, einer relationalen, einer Gruppen- und einer globalen Ebene (Tabelle 12). Monge und Contractor (2003)

sprechen in dieser Hinsicht von multitheoretischen und multi-Ebenen methodentheoretischen Rahmen (Tabelle 12).

Analysen der individuellen Ebene eruieren die Eigenschaften der Knoten. Zentralität ist ein Überbegriff für eine breite Palette von Massen, welche die Wichtigkeit eines Knotens im Netzwerk darstellt. (Brass 1995)

Die relationale Ebene fokussiert auf direkte und indirekte Beziehungen zwischen zwei Knoten. Die wichtigste Art von indirekten Verbindungen sind Pfade. Pfade verlaufen nie mehr als einmal durch denselben Knoten und über dieselbe Kante. Gewicht, Länge und Heterogenität sind die prägenden Eigenschaften der Pfade. Für Kanten ist nur das Gewicht relevant. Die Länge berücksichtigt die Anzahl Kanten entlang des Pfades, die Heterogenität die Anzahl unterschiedlicher Pfade zwischen zwei gegebenen Knoten. (Wasserman und Faust 2008)

Die besondere Herausforderung auf der Gruppenebene ist das Abgrenzen der Knoten. Kohäsive Subgruppen umfassen Knoten, die sehr stark miteinander verbunden sind. Sie weisen starke Beziehungen und eine hohe Dichte innerhalb der Gruppe auf.

Die zweite Möglichkeit der Gruppenbildung basiert auf der Ähnlichkeit der Interaktionsmuster zwischen den Knoten. Knoten, die ähnliche Beziehungen zu ähnlichen Knoten haben, haben dieselbe Position. Direkte und indirekte Beziehungen zwischen Knoten in derselben Position sind nicht notwendig. (Monge und Contractor 2003)

Über die Definition von Rolle und Position herrscht Uneinigkeit (Wasserman und Faust 2008). So bezeichnen Wasserman und Faust (2008) eine Sammlung von Knoten mit ähnlichen Interaktionsmustern als Position und die Kanten zwischen den Knoten und den Positionen als Rollen. Guimerà und Amaral (2005a) sprechen von universellen Rollen und beziehen sich dabei auf Sammlungen von Knoten. Brass (1995) bezeichnet mit dem Begriff Rolle, Knoten mit besonderen Vermittlungseigenschaften im Netzwerk. Die Begrifflichkeit in dieser Arbeit folgt den Bezeichnungen von Guimerà und Amaral (2005a).

Möglich ist auch eine Gruppenbildung aufgrund exogener Attribute. In diesem Zusammenhang spricht man von Partitionen.

Auf globaler Ebene ist die Grösse das prägende Merkmal. Dazu lassen sich aus den einzelnen Massen auf individueller und relationaler Ebene weitere Eigenschaften des Netzwerks ableiten: aus den Kantengewichten die Dichte, aus den Pfadlängen die Konnektivität (oder Erreichbarkeit), aus den Zentralitäten der Grad der Zentralisation. (Brass 1995)

Tabelle 12: Konzepte und Masse der Netzwerkanalyse unterteilt nach der Ebene der Analyse

Analyseebene	Netzwerkanalytische Primitiven	Konzepte und Masse
Individuelle Ebene	Knoten	Zentralität
Relationale Ebene	Kante	Pfad Gewicht Länge Heterogenität
Gruppenebene	Gruppe	Subgruppe Rolle Partition
Globale Ebene	Netzwerk	Dichte Grad der Zentralisation

Die Übereinstimmung der Konzepte und Masse der Netzwerkanalyse (Tabelle 12) mit den Konzepten der metaphorischen Kartierung (Tabelle 11) ist sehr hoch. Mit Methoden der Dimensionsreduktion und der geometrischen Generalisierung können Eigenschaften des Netzwerks besonders hervorgehoben werden.

Geometrie und Netzwerkvisualisierung

Die graphische Darstellung von sozialen Netzwerken ist bereits seit den dreissiger Jahren des 20. Jahrhunderts Teil der Forschung. Das erste Soziogramm, definiert als die graphische Darstellung von interpersonellen Beziehungen, wird von Moreno (1933) entworfen. Heute beschäftigt sich ein eigenständiger, kleiner Forschungsbereich der Informatik mit dem *graph drawing*. Das Visualisieren von sozialen Netzwerken ist ein Teilbereich davon. Die Herausforderung dabei ist das Entwerfen einer optimalen Konfiguration, welche ohne vordefinierte Vorgaben die bedeutungsvollsten, latenten Strukturen hervorhebt (Fleischer und Hirsch 2001).

Die Güte eines Layouts wird mit ästhetischen Kriterien gemessen. Solche Kriterien sind die Minimierung der Kantenkreuzungen, –knicke und –längen, die Maximierung der Winkel zwischen den Kanten, Symmetrie und Orthogonalität (Coleman und Parker 1996; Fleischer und Hirsch 2001; Bennett et al. 2007). Purchase et al. (Purchase 1997; Purchase et al. 1997; Purchase 2000; Ware et al. 2002) bestätigen, dass vor allem die Kantenkreuzungen und die Kantenknicke, aber auch die Symmetrie, die Lesbarkeit der Graphen beeinflussen. Keine empirische Evidenz gibt es für die anderen ästhetischen Kriterien. Durch diese Kriterien wird die Anordnung der Knoten im topologischen Raum bestimmt, dabei werden die Layouts in orthogonal, hierarchisch, planar und kräftebasiert unterschieden (für eine ausführliche Übersicht siehe Di Battista et al. (1994). Empirische Teste deuten auf eine höhere Effektivität der kräftebasierten Layouts hin, der Unterschied zu den anderen Layouts ist jedoch nicht signifikant (Purchase 1998, 2000). Kräftebasierte Algorithmen eignen sich besonders für Graphen, bei denen keine vordefinierten Eigenschaften bekannt sind, wie z.B. soziale Netzwerke (Rodgers 2005). Die Knoten werden nach physikalischen Prinzipien im topologischen Raum angeordnet (Brandes 2001a). Konkret werden stark verbundene Knoten nahe zueinander und schwach verbundenen Knoten weit voneinander platziert (Chen 2004). Kräftebasierte Al-

gorithmen sind weit verbreitet und werden in der Informationsvisualisierung anderen Layouts vorgezogen (Chen 2004), weil sie wegen der Analogie zur physischen Umwelt angeblich intuitiv und einfach zu verstehen sind (Brandes 2001a). Bei den traditionellen kräftebasierten Algorithmen, *spring-embedder* von Eades (1984) und deren Weiterentwicklungen durch Fruchterman und Reingold (1991) sowie Kamada und Kawai (1989), um die wichtigsten zu nennen, ist die Anwendung auf grosse oder dichte Netzwerke problematisch (Chen 2004). Neuere Entwicklungen wie der *graph embedder* (GEM) (Frick et al. 1994) und das *distributed recursive (graph) layout* (DrL/VxOrd-Algorithmus) (Shawn et al. 2011) taugen insbesondere für grössere und komplexere Netzwerke. Kräftebasierte Algorithmen können weiter in gewichtete und ungewichtete unterteilt werden. Gewichtete Algorithmen berücksichtigen nicht nur die Präsenz bzw. Abwesenheit sondern auch die Gewichte der Kanten und sind somit genauer.

Die Geometrie der kräftebasierten Layouts entspricht dem ersten Gesetz der Geographie. Insbesondere die gewichtete Variante ist im Prinzip die mathematische Realisierung von Tobler's Gesetz (erstes Gesetz der Geographie) und somit der räumlichen Autokorrelation. Dadurch konvergieren die netzwerkanalytischen und die metaphorischen Konzepte und realisieren sich in der Visualisierung: die Stärke einer Beziehung wird in der Visualisierung mit der Distanz entlang der Trajektorien wiedergegeben, die Netzwerkdichte ist auch als räumliche Dichte wiederzufinden usw.

3.1.3 Geometrische Generalisierung

Die geometrische Generalisierung ist die graphische Umsetzung der relevanten Information und somit der letzte Schritt des *spatialization*-Prozesses. Anlehnend an Robinson et al. (1978) umfasst die geometrische Generalisierung Symbolisierung und Vereinfachung.

Symbolisierung

Mit der Symbolisierung wird das Ziel verfolgt, die mathematischen und kognitiven Eigenschaften der semantischen Primitiven mittels geometrischer Primitiven (Punkte, Linien und Polygone) graphisch wiederzugeben. Punkte, Linien und Polygone können mit graphischen Variablen verändert werden (Fabrikant und Skupin 2005). Bertin (1974) identifiziert neben der Lage in der Ebene sechs Variablen der dritten Dimension oder Farb-Muster-Variablen: Grösse, Helligkeitswert, Muster, Farbe (Farbton), Richtung und Form. Basierend auf weiteren Quellen präsentiert MacEachren (1994) eine erweiterte Taxonomie. Zusätzlich zu den sieben Variablen von Bertin umfasst diese Syntaktik Schärfe, Auflösung, Transparenz, Farbsättigung und Anordnung.

Die Lage der Knoten und Kanten wird bereits mit dem Dimensionsreduktionsverfahren bestimmt, die Farb-Muster-Variablen lassen sich variieren, um weitere Attribute der semantischen Primitiven hervorzuheben.

Graphische Variablen werden durch ihre Gliederungsstufe, welche von der Wahrnehmbarkeit gegeben wird, definiert. Bertin (1974) identifiziert vier Stufen der Wahrnehm-

barkeit: assoziativ, selektiv, geordnet und quantitativ. Je mehr Eigenarten der Wahrnehmbarkeit eine Variable besitzt, desto höher ist ihre Gliederungsstufe (Bertin 1974).

Bei assoziativen Variablen wird die Variation derselben übersehen bzw. nivelliert und die Elemente werden aufgrund anderer Variablen als Gruppe wahrgenommen. Dissoziative Variablen beherrschen hingegen die Wahrnehmung jeder Kombination von Variablen; Gruppen werden spontan aufgrund der betrachteten Variablen gebildet. Bei dissoziativen Variablen ändert sich mit dem Variieren der Kategorien die Sichtbarkeit der Elemente. Dies lässt sich gut am Beispiel der dissoziativen Variable Grösse erklären: grosse Kartenelemente sind besser sichtbar als kleine. Gemäss Bertin (1974) sind Lage, Farbton, Muster, Orientierung und Form assoziativ, Grösse und Helligkeitswert hingegen dissoziativ. MacEachren (1995) spricht in Bezug auf die assoziativen Eigenschaften von *visual levels*. Dies unterstreicht vor allem die unterschiedlichen Niveaus, die Hierarchie, welche sich bei dissoziativen Variablen bilden. Bei der Variable Muster herrscht Uneinigkeit: gemäss Bertin (1974) ist das Muster assoziativ und gemäss MacEachren (1995) hingegen eher dissoziativ. In der Wahrnehmungspsychologie spricht man von *divided attention* (Pomerantz und Schwaizberg 1975).

Von selektiver Wahrnehmung spricht man bei der Möglichkeit, eine bestimmte Kategorie zu isolieren, Elemente nach dieser zu gruppieren und dabei die anderen Variablen zu ignorieren (Bertin 1974). MacEachren (1995) spricht dabei von *visual isolation* und Pomerantz und Schwaizberg (1975) von *selectiv attention*. Die Form und in den flächenhaften Darstellungen die Richtung sind die einzigen Variable nach Bertin welche nicht selektiv sind (Bertin 1974).

Bei geordneten Variablen lässt sich spontan die Ordnung der Kategorien feststellen, deren quantitative Unterschiede lassen sich jedoch nicht intuitiv bestimmen. Die präzisen numerischen Unterschiede zwischen den Kategorien lassen sich nur bei quantitativen Variablen erkennen (Bertin 1974). Ware (2004) spricht von monotonen Variablen. Muster und Helligkeit sind geordnete bzw. monotone Variablen. Lage und Grösse gelten sowohl als geordnete als auch als quantitative Variablen (Bertin 1974).

Die Wahl der Variablen für die Visualisierung von Daten ist optimal, wenn das Skalenniveau der Daten und die Gliederungsstufen der Variablen übereinstimmen. Grundsätzlich gilt, dass quantitative Variablen insbesondere für metrische, geordnete Variablen für ordinale und die restlichen Variablen für qualitative Daten geeignet sind.

Die Assoziativität und die Selektivität bestimmen auch die Eignung der Variablen für uni- und multidimensionale Darstellungen. In dieser Hinsicht spricht man von integralen und trennbaren Dimensionen bzw. Variablen. Trennbare Variablen werden von den Lesenden einzeln wahrgenommen, integrale Variablen sind hingegen holistisch, sie wirken als Ganzes. Typische trennbare Variablenkombinationen sind Lage und Grösse, Form und Farbe. Integrale Kombination ist hingegen Grösse und Form. (Ware 2004)

Wenige empirische Studien haben sich mit der Wirksamkeit der Variablen beschäftigt. In einer Metastudie identifizieren Wolfe und Horowitz (2004) Farbe, Orientierung und

Grösse als Variablen, die in statischen Darstellungen die Aufmerksamkeit mit Sicherheit lenken. Bei Helligkeit und Form ist es wahrscheinlich, dass sie die Aufmerksamkeit lenken. Die Wirksamkeit von Farbe, Grösse, Helligkeit, Form und Orientierung wird von Garlandini und Fabrikant (2009) empirisch geprüft. Grösse hat die stärkste Wirkung auf die Aufmerksamkeit, gefolgt von Farbe und Helligkeit. Am schwächsten wird die Aufmerksamkeit von der Orientierung gelenkt. Nach Bertin (1974) ist die Grösse die einzige Variable, welche sowohl selektiv, als auch geordnet, quantitativ und dissoziativ ist, dabei erstaunt die starke Wirkung auf die Aufmerksamkeit nicht. Anders sieht es bei der Helligkeit aus, diese schneidet unter den Erwartungen ab (Garlandini und Fabrikant 2009).

Vereinfachung

Mit der Vereinfachung wird ausschliesslich die Geometrie verändert, um die Lesbarkeit zu erhöhen ohne damit zu viel an Information zu verlieren. McMaster (1989) entwickelt eine detailliertere Taxonomie der Operatoren als Robinson et al. (1978) und unterscheidet die Operationen nach Datenmodell (Raster vs. Vektor) und nach Geometrie (Punkt, Linie und Fläche). In der Netzwerkvisualisierung kommen alle geometrischen Primitiven vor, jedoch ausschliesslich als Vektoren. Das Zusammenfassen ist der wichtigste Operator im *spatialization*-Prozess: Naheliegende Punkte, parallelverlaufende Linien oder angrenzende Flächen werden zu einem Punkt (*aggregation*), einer Linie (*merging*) bzw. einer Fläche (*amalgamation*) zusammengefasst (Regnauld und McMaster 2007). Das Verdrängen ist ein weiterer Operator der geometrischen Generalisierung, der jedoch nicht bei der *spatialization* angewendet werden sollte. Die Lage drückt die topologische Beziehung zu den anderen Elementen im Raum aus und ist somit die wichtigste Eigenschaft jedes Elementes. Eine Verdrängung würde diese verändern und beeinflusst nach Fabrikant et al. (2004) die Interpretation der Visualisierung. Zusammenfallen oder der Geometrietypenwechsel bezieht sich auf das Ändern der geometrischen Primitiven, insbesondere auf das Ersetzen von Flächen durch Punkte (Regnauld und McMaster 2007). Im *spatialization*-Prozess ist jedoch auch die entgegengesetzte Änderung vorstellbar. Wenn es der Massstab erlaubt, verstärkt das Ersetzen von Punkten durch Flächen das Container-Image-Schema und somit die Wahrnehmung von einem Inneren und einem Äusseren (Salvini et al. 2011).

3.1.4 Fazit: Grammatik der Netzwerkvisualisierung

In Anlehnung an Ware (2004) lässt sich eine Grammatik der graphischen Darstellungen definieren, so auch von Netzwerkvisualisierungen (Tabelle 13). Die semantischen Primitiven werden mittels geeigneten geometrischen Primitiven dargestellt und graphische Variable heben besondere Konzepte hervor.

Tabelle 13: Grammatik der graphischen Darstellung

Semantische Primitiven	Geometrische Primitiven	Graphische Variable	Netzwerkanalytische Konzepte
Lokus	Punkt / Polygon	Grösse	Zentralität
		Helligkeit	ev. Zentralität
		Farbton	Gruppe
		Farbton + Helligkeit	Hierarchie
		Farbton + Sättigung	Hierarchie
		Form	Rolle
Trajektorie	Linie	Grösse	Gewicht
		Helligkeit	Gewicht
		Farbton	Typ
Aggregat	Polygon / Punkt	Grösse	Grösse
		Helligkeit	Dichte
			Konnektivität
			Grad der Zentralisation
		Farbton	Typ (z.B. Identität)
Grenze (=Polygonkontur)	Linie	Textur	Diskontinuität

Loci werden in Abhängigkeit vom Massstab mit Punkten oder Polygonen dargestellt. Die Zentralität als wichtigste Eigenschaft der Knoten lässt sich mittels der graphischen Variable Grösse darstellen. Grösse besitzt alle Gliederungsstufen und ermöglicht somit die intuitive Wahrnehmung des metrischen Skalenniveaus. Ein zweites Zentralitätsmass lässt sich über Variationen der Helligkeiten darstellen, oder die Grösse wird in zwei Komponenten (Breite vs. Länge) unterteilt. Durch die Helligkeit geht das metrische Niveau der Zentralität verloren. Durch die Variation der Breite und Länge (integrale Variablen) werden die zwei Zentralitätsmasse jedoch als eine Einheit und nicht mehr als zwei Dimensionen wahrgenommen.

Die Zugehörigkeit eines Lokus zu einer Gruppe (Partition, Rolle oder Subgruppe) lässt sich über den Farbton zeigen. Der Farbton ist assoziativ und selektiv, was das nominale Skalenniveau der Gruppenzugehörigkeit optimal wiedergibt. Die Gruppenzugehörigkeit wird in der Tendenz bereits mit der Lage suggeriert: Knoten in der Nähe oder gar im selben Cluster gehören zur selben Gruppe. Die redundante Darstellung mittels Farbton stärkt die Wahrnehmung (Fabrikant et al. 2006). Die Stärkung ergibt sich durch die Koppelung von zwei Gsgesetzen: das Gesetz der Nähe wird durch das Gesetz der Ähnlichkeit verstärkt. Die hierarchische Organisation der Gruppen lässt sich durch die Kombination von Farbton und Helligkeit oder Farbton und Sättigung wiedergeben. Die Kategorien der obersten hierarchischen Ebene werden mittels unterschiedlichen Farbtönen, die folgenden intragruppen Unterteilungen mittels Variation der Helligkeit oder Sättigung wiedergegeben (Fabrikant et al. 2006).

Linien geben die Trajektorien wieder. Mit der Variation der Grösse wird das Gewicht der Kanten am salientesten wiedergegeben (Fabrikant et al. 2004), alternativ wäre eine Darstellung mit der Helligkeit möglich. Eine Kombination der beiden Variablen stärkt durch die Redundanz die Wahrnehmung. Falls die Trajektorien unterschiedlichen Arten von Relationen entsprechen, lassen sie sich mit unterschiedlichen Farbtönen darstellen.

Die Semantik der Aggregate wird je nach Massstab mit Punkten oder Polygonen erfasst. Den Grundregeln der Kartographie zufolge, lässt sich deren Grösse als absolute Variable mit der graphischen Variable Grösse darstellen. Die Dichte, die Konnektivität und die Zentralisation, drei relative Variablen, lassen sich hingegen mit der Helligkeit wiedergeben. Die Typologie eines Aggregats, was zum Beispiel auch seiner Identität entsprechen könnte, lässt sich wiederum mittels Farbton darstellen. Die Identität eines Objektes wird meistens mit der Beschriftung aufgezeigt. Da Aggregate in der Regel das Resultat von Klassifikationen sind, ist deren Anzahl gering. Dadurch ist eine Hervorhebung der Identität mit dem Farbton denkbar. Bei Loci wäre dies hingegen aufgrund ihrer hohen Anzahl undenkbar.

Fabrikant et al. (2006) finden keinen Einfluss der Darstellung der Polygongrenzen auf die Wahrnehmung. In diesen empirischen Studien sind sie konstant, also ohne Variation der Farb-Muster-Variablen dargestellt. Eine Darstellung der Konturen ist somit nur sinnvoll, wenn weitere Informationen durch deren Variation vermittelt werden. In Salvini et al. (2011) wird die Textur der Linien geändert, um Diskontinuitäten bzw. fehlende Interaktionen zwischen angrenzenden Knoten hervorzuheben.

3.2 Methodische Umsetzung

Die folgenden zwei methodischen Abläufe setzen die Kernelemente des *spatialization frameworks* praktisch um. Der erste Ablauf erzeugt eine Punkt-Linien- bzw. eine Netzwerkvisualisierung und der zweite eine *self-organizing map* (SOM). Dabei liegt der Fokus auf den Dimensionsreduktionsverfahren und auf der geometrischen Generalisierung. Parallel dazu wird das Vorgehen in Bezug auf die metaphorische Kartierung gesetzt. Das Vorgehen ist auf die Visualisierung und Analyse von jeglichen Netzwerken anwendbar.

Punkt-Linien-Visualisierung

Der Ausgangsdatensatz ist eine Ähnlichkeits- oder eine Verbundenheitsmatrix, welche die Struktur des Netzwerks in sich birgt. Die Matrixwerte sind die Kantengewichte des Netzwerkes.

Der erste Aspekt, der für einen visuell-analytischen Ansatz zu berücksichtigen ist, ist die Dimensionsreduktion bzw. die Verräumlichung des multidimensionalen Datensatzes. Wie oben erläutert ermöglichen kräftebasierte Layoutalgorithmen die Verräumlichung der Netzwerke unter Einhaltung der Distanz-Ähnlichkeitsmetapher bzw. der Distanz-Verbundenheitsmetapher. Von den unzähligen Algorithmen wird der DrL/VxOrd-

Algorithmus gewählt (für Details zum Algorithmus Shawn et al. 2011) und der GEM-Algorithmus (für Details zum Algorithmus Frick et al. 1994).

Der DrL/VxOrd-Algorithmus ist für grosse bis sehr grosse Netzwerke entwickelt (NWB Team 2011) und ist im Network Workbench Tool (NWB) (NWB Team 2006) integriert. Die Evaluationen von Klavans und Boyack (2006a, 2006b; Boyack et al. 2005) ergeben, dass der DrL/VxOrd-Algorithmus die latenten Strukturen der Daten genau wiedergibt. Untersucht werden die lokale und die globale Genauigkeit der Visualisierung. Beide sind hoch. Mit der lokalen Genauigkeit untersucht man, ob die Distanz eines Objektes zu seinen Nachbarn im selben Cluster korrekt ist (Klavans und Boyack 2006a). Die globale Genauigkeit untersucht die Zusammensetzung und die Platzierung der Cluster (Klavans und Boyack 2006b). Die Genauigkeit ist maximal, wenn der Input für den DrL/VxOrd-Algorithmus nicht aus der gesamten Matrix besteht, sondern für jeden Knoten nur die 15 wichtigsten Beziehungen gewählt werden (Klavans und Boyack (2006a). Dadurch wird vermieden, dass sich ein undifferenziertes Layout ergibt, welches vollständig verbunden erscheint und dessen Muster und Strukturen nicht ersichtlich sind (Shawn et al. 2011). Eine systematische Evaluation dieses Grenzwertes (15 Kanten pro Knoten) fehlt bisher in der Literatur und wird in dieser Arbeit vorgenommen (Kapitel 6.2).

Im DrL/VxOrd-Algorithmus ist eine zusätzliche Kantenfilterung, das *edge cutting*, integriert. Das Ziel ist wiederum, die Lesbarkeit der Visualisierungen zu erhöhen. Durch die Filterung werden Kanten, die während dem Dimensionsreduktionsverfahren wegen ihrer Länge ein bestimmtes Gewicht-Länge-Verhältnis unterschreiten, gelöscht. Die Stärke der Filterung wird mittels eines Parameters, welcher zwischen Null, keine Filterung, und eins, starke Filterung, variiert, gesteuert. Bei einem hohen Wert tendiert die Verteilung der Knoten im Raum zu einer stärkeren Clusterbildung. Mit 0.8 wird eine optimale Verteilung erreicht. (Shawn et al. 2011; Klavans und Boyack 2006a)

Der ungewichtete GEM-Algorithmus (Frick et al. 1994) ist auch in NWB (NWB Team 2006) implementiert und ist auch für kleinere Netzwerke geeignet. Er hat den Nachteil, dass die Kantengewichte bei der Dimensionsreduktion unberücksichtigt bleiben. Die Kanten werden nur in binärer Form betrachtet: vorhanden, nicht-vorhanden. Die Distanzmetrik im topologischen Raum ist somit weniger genau: verbundene Knoten werden nahe zueinander platziert, nicht-verbundene werden auseinander gedrückt. Die Stärke der Verbindungen findet jedoch keinen Gegenwert in der Distanz. Der entscheidende Vorteil ist die gute Lesbarkeit der erzeugten Visualisierungen, welche sich bei wenig dichten Netzwerken mit der Minimierung der Kantenkreuzungen ergibt. Da im GEM-Algorithmus keine Kantenfilterung enthalten ist, wird das Netzwerk vorweg mit der *pathfinder network scaling* Methode (Dearholt und Schvanenveldt 1990) reduziert.

Die *pathfinder network scaling* Methode (Dearholt und Schvanenveldt 1990) ist in der Informationsvisualisierung weit verbreitet (Börner et al. 2003) und vereinfacht komplexe Netzwerke indem wenige, wichtige Kanten extrahiert werden, welche dem Muster des Netzwerks zugrunde liegen (Chen 1999). Das *pathfinder network scaling* basiert auf dem

mathematischen Prinzip der Dreiecksungleichung. Diese besagt, dass eine Seite eines Dreiecks nur kürzer oder maximal gleich lang sein kann wie die Summe der anderen zwei Seiten. Auf das Netzwerk übertragen bedeutet dies, dass eine Kante zwischen zwei Knoten nur erhalten bleibt, wenn kein kürzerer indirekter Pfad existiert. Zwei Parameter bestimmen das Resultat die *pathfinder network*-Vereinfachung: der r -Parameter bestimmt die Minkowski-Metrik für die Funktion mit der aus den Kantengewichten die Pfadlänge berechnet wird; der q -Parameter bestimmt die maximale Anzahl Links, aus denen der indirekte Pfad besteht (Chen 1999; Dearholt und Schvanenveldt 1990). Bei der Implementierung in NWB (NWB Team 2006) ist q unveränderlich und auf $N-1$ (mit N =Anzahl Knoten) gesetzt, was für Visualisierungsanliegen besonders geeignet ist (Chen 1999). Der r -Parameter kann hingegen variiert werden. Für $r=1$ entspricht die Distanz der einfachen Summe der Gewichte, für $r=2$ der euklidischen Distanz und für $r=\infty$ entspricht die Distanz dem maximalen Gewicht auf dem Pfad (Dearholt und Schvanenveldt 1990). Je grösser r ist, desto weniger Kanten bleiben im Netzwerk erhalten (Börner et al. 2003).

Bei der Visualisierung mit dem GEM-Algorithmus wird, wegen der optimalen Lösungen bei wenig dichten Matrizen, der MST ($r=\infty$) berechnet und aufgrund dessen das Layout generiert.

Beide Dimensionsreduktionsverfahren ergeben in erster Linie eine Punkt-Visualisierung. Jeder Punkt stellt einen Knoten dar. Obwohl teilweise klare Strukturen zum Vorschein kommen, sind die Lesbarkeit und die Salienz der Visualisierung bescheiden. Wegen der hohen Dichte überlagern sich die Punkte teilweise und verdecken sich gegenseitig. Aus diesem Grund werden im nächsten Schritt die Punkt-Visualisierungen mit etablierten kartographischen Methoden in ArcMap (ESRI 2011) im Hinblick auf Salienz und Lesbarkeit optimiert (Fabrikant und Salvini 2011). Zwei Operationen der geometrischen Generalisierung werden dabei angewendet: Aggregation und Zusammenfallen. Mit der Aggregation werden Cluster von naheliegenden Punkten zu Polygonen zusammengefasst. Diese Polygone werden wiederum durch Punkte ersetzt (Zusammenfallen), deren Koordinaten dem Zentroid des Polygons entsprechen.

Die generierten Informationsräume entsprechen einem diskreten und abrupten Datenmodell (MacEachren 1995). Diskret weil die Entitäten (Knoten) nicht kontinuierlich über den ganz Raum verteilt sind, sondern nur an gewissen fokalen Punkten vorkommen und abrupt, weil die Grenzen zwischen den Entitäten klar und scharf gezogen sind. Aus semantischer Perspektive entsprechen die Punkte in der ungeneralisierten Darstellung den Loci: die kleinsten Einheiten mit eindeutiger und aussagekräftiger Lage. Durch die Generalisierung werden die Loci zu Aggregaten zusammengefasst. Das Aggregationskriterium ist die ähnliche Lage, bzw. die enge Verbundenheit.

Die intrinsische Essenz einer Netzwerkvisualisierung sind die Verbindungen. Semantisch gesehen spricht man von Trajektorien. Da die Kanten in einem Netzwerk oft unzählig sind, führt die Darstellung von allen Kanten zu einer graphischen Überladung. Von den unterschiedlichen Filtermethoden hat sich das *pathfinder network scaling* etabliert

(Börner et al. 2003). Der Algorithmus wird nicht auf das ursprüngliche Netzwerk angewendet, sondern auf das generalisierte. Das *pathfinder network* welches sich zwischen den aggregierten Punkten aufspannt, wird mit $r=\infty$ berechnet und dargestellt. Da das GEM-Layout die Distanz-Ähnlichkeitsmetapher nicht optimal wiedergibt, ist es unentbehrlich, die Stärke der Verbindungen mit Helligkeit und Grösse der Kanten hervorzuheben (Fabrikant et al. 2004). Die Kanten werden analog dazu auch im DrL/VxOrd-Layout dargestellt.

In dieser rohen Punkt-Linien-Darstellung ist die Lage der Knoten die einzige verwendete graphische Variable. Attribute der Trajektorien und der Aggregate können nach Tabelle 13 mit weiteren Variablen hervorgehoben werden. Farbton, Helligkeit, Grösse und Form sind dabei die Variablen mit der breitesten Anwendung. Die Farbwahl folgt den Richtlinien nach Brewer (2005, 2006). Aus kartographischer Perspektive ist der Einsatz von Diagrammen, insbesondere Kreisdiagrammen, eine aussagekräftige Darstellungsart für die Aggregate. Dadurch lassen sich deren Grösse und Vielfalt hervorheben.

SOM

Netzwerkvisualisierungen sind sehr aussagekräftige Informationsvermittler, wenn Verbundenheit im Zentrum des Interesses steht: welche Knoten, oder Gruppen von Knoten stehen in Bezug zueinander? Aus Lakoff und Johnson's (1980) Perspektive sind Netzwerkvisualisierungen die Umsetzung des Link- und des Pfad-Image-Schemas. Das Container-Image-Schema ist jedoch schwach umgesetzt. Das Enthalten oder Enthalten-sein kommt nicht explizit zum Ausdruck. Das Container-Schema ist z.B. bei Fragestellungen wie "Zu welcher Funktion gehört ein Akteur?" durchaus von Bedeutung. Diese Fragestellungen beruhen grundsätzlich auf bimodalen Netzwerken, wie z.B. Akteur-Funktion-Netzwerke. In diesen Fällen bietet sich die SOM als Visualisierungsmethode an.

Mittels SOM lässt sich eine grosse Menge von multidimensionalen Daten organisieren und visualisieren (Börner et al. 2003). SOM gehören zu den artifiziellen, neuronalen Netzwerkmethoden. Eine Menge an multidimensionalen Inputdaten werden in einem zweidimensionalen topologischen Raum, welcher aus hexagonalen Neuronen besteht, projiziert (für mathematische Details siehe Kohonen 2001). Diese Neuronen nehmen die gleichen Attribute wie die Inputdaten an (Skupin und Agrawal 2008).

SOM sind sowohl eine Clustertechnik als auch ein Dimensionsreduktionsverfahren. Die Dimensionsreduktion ist durch das Projizieren der n-dimensionalen Daten in den zweidimensionalen Raum gegeben. Wenn der topologische Raum mit vielen Neuronen angelegt wird, in dem tendenziell jedes Neuron einem Inputdatensatz entspricht, wird dieses Verfahren verstärkt. Wenn hingegen wenige Neuronen angelegt werden und diese mehrere Inputdatensätze beinhalten, wird das Clustern verstärkt. (Skupin und Agrawal 2008)

SOM bestehen aus zwei unterschiedlichen Ebenen: dem Komponentenraum und der *best-matching unit* (BMU). Der Komponentenraum umfasst eine definierte Anzahl Neuronen. Die Attribute dieser Neuronen werden aus den Inputdaten ermittelt. Die BMU

besteht aus allen Inputdaten; jeder Inputdatensatz entspricht einem Punkt. Diese werden in den Komponentenraum projiziert. Jeder Punkt der BMU ist im Neuron enthalten, welches seine Eigenschaften am besten abbildet.

Für die Berechnung der SOM wird das SOM Analyst Tool (Lacayo-Emeri 2011b) in ArcGIS verwendet. Der Input besteht aus einer bimodalen Matrix, wobei im Voraus zu bestimmen ist, welcher Modus der Merkmalsträger und welcher das Merkmal ist. Der Arbeitsablauf richtet sich nach Lacayo-Emeri (2011a). Die Merkmale werden zu Beginn zwischen null und eins normalisiert. Die Grösse des Komponentenraums und die Trainingslänge werden in Abhängigkeit der Datenmenge, des Zweckes (Klassierung bzw. Dimensionsreduktion) und der Rechenkapazität bestimmt.

Auf konzeptioneller Ebene entspricht die Visualisierung nur im Ansatz der Distanz-Ähnlichkeitsmetapher: ähnliche Neuronen liegen nahe beieinander und ähnliche Inputdaten werden in gleiche oder benachbarte Neuronen projiziert. Die Ähnlichkeit zwischen benachbarten Neuronen ist jedoch nicht gleich über die ganze SOM, obwohl die Distanz konstant ist. Der Distanzfehler wird in der U-Matrix gespeichert (Lacayo-Emeri 2011b): diese gibt die Ähnlichkeiten benachbarter Neuronen wieder. Mit der U-Matrix lässt sich der topologische Raum in Bezug auf die Distanz-Ähnlichkeitsmetapher optimieren. Die SOM wird aufgrund der U-Matrix-Werte mit Scape Toad (Andrieu et al. 2008) in ein Kartogramm umgerechnet. Dabei werden Neuronen mit grossen U-Werten und somit geringerer Ähnlichkeit zu den Nachbarn grösser, sehr ähnliche Neuronen werden kleiner. Die Veränderung der Grösse verändert auch die Distanz zwischen den Zentrumspunkten der Neuronen; sehr ähnliche Neuronen rücken zusammen, unähnliche rücken auseinander.

Der vorliegende topologische Raum öffnet unterschiedliche Visualisierungsmöglichkeiten:

- **SOM als Symbolkarte** (z.B. Abbildung 30): Die verzerrten Neuronen werden durch Punkte in deren Schwerpunkten ersetzt. Aus geometrischer Perspektive wird die Distanz-Ähnlichkeitsmetapher optimiert. Jeder Punkt wird zum Träger der Merkmale des Neurons im Komponentenraum. Die Variablen lassen sich einzeln mit Farb-Muster-Variationen darstellen. Idealerweise könnte die Grösse der Punkte aufgrund der Anzahl Inputdatensatz, die in einem Neuron enthalten sind, variiert werden. Dadurch wäre klar ersichtlich, welche Bereiche des topologischen Raumes tatsächlich mit Daten besetzt sind und welche nicht. Die Variation der Grösse ist jedoch aus Platzgründen nicht immer möglich. Alternativ dazu kann diese Information auch binär vermittelt werden: Punkte werden mit einheitlicher Grösse dargestellt, wenn das Neuron besetzt ist, Punkte werden nicht dargestellt, wenn sie unbesetzt sind. Durch das Betrachten des Verlaufs der verschiedenen Variablen über den Raum lässt sich deren Zusammenhang ermitteln: gleicher Verlauf entspricht einer hoher Korrelation zwischen den Variablen.

Es ist auch möglich alle Merkmale gleichzeitig zu visualisieren. Dabei sind die Variablen im Voraus mit der Min-Max-Methode zu normalisieren. Jedem Punkt wird dann der Wert der Variable mit der maximalen Ausprägung zugewiesen. Mit der Farbtonvariation wird die Variable hervorgehoben, mit der Helligkeitsvariation die Ausprägung derselben.

Das Datenmodell entspricht dem diskreten semiabrupten Modell (MacEachren 1995). Die einzelnen Punkte entsprechen aus semantischer Perspektive den Aggregaten. Wobei, insbesondere wenn die Grösse des Komponentenraumes der Anzahl Inputdatensätze entspricht, viele Neuronen genau einen Merkmalsträger enthalten und somit einem Lokus entsprechen. Mehrere Punkte, welche durch eine gemeinsame Farbgebung gekennzeichnet sind, bilden wiederum ein Aggregat.

- **SOM als Container-Raum** (z.B. Abbildung 45): Wie einführend dargelegt, ermöglichen SOM die Hervorhebung der Container-Metapher. Dies erfolgt dank einem Zwei-Ebenen-System. Die untere Ebene besteht aus dem verzerrten Komponentenraum, in dem die Neuronen als Polygone dargestellt werden. Die obere Ebene besteht aus der BMU, indem jeder Inputdatensatz durch einen Punkt dargestellt wird. Jeder Punkt ist somit in einem Neuron enthalten.

Selten lässt sich jedes Neuron und vor allem die BMU unverändert darstellen. Mehrere Punkte der BMU, welche im gleichen Neuron liegen, werden zu einem einzigen Punkt zusammengefasst (Punkt-Aggregation). Analog dazu lassen sich angrenzende und gleiche bzw. ähnliche Neuronen zu Regionen aggregieren.

Besonders geeignet ist der Container-Raum, wenn auf der BMU andere Variablen als auf dem Komponentenraum dargestellt werden. Dies lässt sich am Besten anhand eines Beispiels (Abbildung 45) erläutern. Die Inputmatrix setzt Städte in Bezug zu Funktionen und anhand dieser wird der Komponentenraum berechnet. Die unterschiedlichen Funktionen lassen sich nun, wie oben erwähnt, im Komponentenraum mit unterschiedlichen Farbtönen und Helligkeiten visualisieren. Es bilden sich somit Regionen mit gleicher Farbgebung. In der BMU wird jede Stadt durch einen Punkt dargestellt. Auf diesen Punkt lassen sich exogene Variablen darstellen, z.B. die Zentralität einer Stadt. Damit wird salient dargestellt, welche Stadt (Punkt) zu welchem funktionalen Raum (Farbe der Polygone) gehört und wie zentral (Grösse des Punkts) die Stadt ist. Wenn die BMU geometrisch generalisiert wird (Punkt-Aggregation), werden die Attribute dementsprechend in aggregierter Form neu berechnet.

Das Datenmodell entspricht dem kontinuierlichen semiabrupten Modell (MacEachren 1995). Kontinuierlich, weil die Neuronen ohne Unterbruch über den ganzen Raum verteilt und semi-abrupt, weil die Grenzen zwischen den Neuronen klar und scharf gezogen sind, die Wertevariation zwischen benachbarten Neuronen aber klein sind. Neuronen und Gruppen von Neuronen im Kompo-

nentenraum entsprechen den Aggregaten auf unterschiedlichen Massstabsebenen. Die Aggregatsmetapher wird durch die flächenhafte Darstellung verstärkt. Die Punkte der BMU entsprechen wiederum je nach Generalisierung einzelnen Loci oder Aggregaten.

4 Datengrundlage

Dieses Kapitel stellt die Datengrundlage Wikipedia vor. Dabei werden neben der Datenstruktur die wichtigsten Chancen und Risiken der wissenschaftlichen Verwendung diskutiert und die bestehenden wissenschaftlichen Ansätze zum Thema skizziert.

Die Wissensproduktion und –verbreitung hat sich in den letzten Jahren markant verändert. Dieser Wandel zeigt sich sehr klar im Bereich der topographischen Karten. Staatliche Ämter, in der Schweiz das Bundesamt für Landestopographie, erheben Geodaten und produzieren topographische Karten nach präzisen Standards und basierend auf dem Fachwissen der Mitarbeitende. Dieser *top-down* Ansatz bleibt im Wesentlichen unverändert, wird jedoch durch *bottom-up* Produkte ergänzt und konkurriert. Ein bekanntes Beispiel ist *OpenStreetMap*. Bei *OpenStreetMap* können sich alle bei der Datenerhebung sowie bei der graphischen Umsetzung beteiligen (OpenStreetMap contributors 2011b). 2011 zählt *OpenStreetMap* über 40'000 Mitglieder (OpenStreetMap contributors 2011a). Das Fachwissen der Mitglieder bleibt ungewiss, im Gegenzug ergeben sich Vorteile durch die hohe Anzahl der Beteiligten. Howe (2006) beschreibt diesen *bottom-up* Prozess als „*outsourcing [a function] to undefined (and generally large) network of people in the form of an open call*“ und nennt ihn *crowdsourcing*. Das Potential dieses Ansatzes liegt einerseits in der Masse, in der hohen Anzahl an Beteiligten, welche sich meistens freiwillig mit dem Projekt beschäftigen, andererseits in der Netzwerkorganisation, welche das Teilen und Aufbauen von und das Diskutieren über Wissen ermöglicht. Das *World Wide Web*, insbesondere das Web 2.0 mit interaktiven und kollaborativen Elementen, ist Voraussetzung für ein effektives *crowdsourcing*. Wegen der grossen Relevanz des *crowdsourcing* in unserer Ge-

sellschaft hat *Time* (Grossman 2006) *You* zur Person des Jahres 2006 gekürt (Abbildung 11).

Abbildung 11: Titelseite der Dezember 2006 Ausgabe von *Time*



Das Resultat eines gigantischen *crowdsourcing*-Projekts ist die grösste Sammlung an frei zugänglichem Wissen (Zech et al. 2007): die Online Enzyklopädie Wikipedia. Das Projekt wurde 2001 von Jimmy Wales und Larry Sanger ins Leben gerufen (Wikipedia contributors 2011d). 2011 umfasst die Enzyklopädie 281 Versionen in unterschiedlichen Sprachen und insgesamt ca. 19 Millionen Artikel. Mehr als ein Drittel entfällt auf die drei Versionen mit mehr als 1'000'000 Artikeln (Englisch, Deutsch, Französisch). Knapp 30 Millionen AutorInnen schreiben freiwillig die Inhalte der Enzyklopädie, welche von knapp 5'000 AdministratorInnen beaufsichtigt werden. (Wikipedia contributors 2011c)

Gemäss den Wikipedia Regeln wird ausschliesslich enzyklopädisch relevantes Wissen in die Enzyklopädie aufgenommen. Die Relevanz wird wie folgt definiert: „*Die Entscheidung für oder gegen die Aufnahme in eine Enzyklopädie richtet sich auch nach der Frage, ob Personen, Ereignisse oder Themen mit aktuell breiter Öffentlichkeitswirkung nach sinnvollem Ermessen auch Zeit überdauernd von Bedeutung sein werden*“ (Wikipedia contributors 2011e). Trotz der regulierenden Eingriffe der AdministratorInnen kann Wikipedia als ein *crowdsourcing* Projekt betrachtet werden, welches aus der freien Zusammenarbeit auf globaler Ebene entsteht.

Wegen der offenen Natur von Wikipedia wird deren Inhalt oft kontrovers diskutiert. Einerseits werden der Mehrwert und die Genauigkeit, welche aus der kollektiven Zusammenarbeit entstehen (Raymond 1997; Flanagin und Metzger 2008) anderseits werden die Risiken in Bezug auf die Reliabilität und Genauigkeit der Texte (Denning et al. 2005) hervorgehoben. Denning et al. (2005) identifizieren sechs zentrale Risiken, denen man auf jeder Wikipedia-Seite begegnen kann:

- Genauigkeit der Einträge. Es bleibt ungewiss, ob die Information zum betrachteten Artikel genau oder ungenau ist. Giles (2005) ist dieser Frage nachgegangen und hat wissenschaftliche Artikel in Britannica und Wikipedia mit einem Peer-Review-Verfahren verglichen und dabei herausgefunden, dass die beiden Enzyklopädien ähnlich genau bzw. ungenau sind. Flanagin und Metzger (2008) haben einen direkten Zusammenhang zwischen Genauigkeit und Glaubwürdigkeit der Einträge einerseits und der Anzahl AutorInnen pro Artikel anderseits gefunden.
- Volatilität der Einträge. Die vorgenommenen Änderungen der AutorInnen werden gleich vollzogen (Wikipedia contributors 2011d). Insbesondere kontroverse Artikel können mehrmals auch in kurzer Zeit geändert werden, was zu hoher Aktualität (Zech et al. 2007), aber auch zu hoher Instabilität führt.
- Schreibgründe, Fachwissen und Quellen. Die Richtlinien für AutorInnen sehen vor, dass Beiträge aus einem neutralen Standpunkt geschrieben und mit verifizierbaren Quellen untermauert werden müssen (Wikipedia contributors 2011d). Denning et al. (2005) sehen die ungewissen Beweggründe für das Schreiben in Wikipedia, das heterogene Fachwissen sowie die beschränkte Anzahl Quellen weiterhin als ein offenes Problem. Elwood (2010) spricht in dieser Hinsicht von semantischer Heterogenität: die Beiträge unterscheiden sich hinsichtlich Form und Inhalt stark voneinander.
- Thematische Abdeckung. Da die Artikel auf freiwilliger Basis entstehen, widerspiegeln die abgedeckten Themen die Interessen der Autorenschaft. Wikipedia-AutorInnen sind vor allem junge, gebildete Männer ohne Kinder (Glott et al. 2010), die die Sprache der Enzyklopädie beherrschen. Die beschränkte Vielfalt der Autorenschaft zeigt sich verstärkt im Inhalt, weil Leute tendenziell auch über räumlich und zeitlich nahe Ereignisse und Gegebenheiten schreiben (Flanagin und Metzger 2008). Dies führt zu thematischen, zeitlichen und räumlichen Verzerrungen. Die thematische Verzerrung haben Kittur et al. (2009) beschrieben, indem sie die Artikel der englischen Wikipedia elf Themen zugewiesen haben. Die meisten davon (30%) gehören zum Thema Kultur und Kunst und die wenigsten (1%) zum Thema Philosophie und Denken.

Struktur von Wikipedia

Wikipedia besteht aus struktureller Perspektive aus zwei kommunizierenden Systemen: Artikelsystem und Kategoriensystem. Das Artikelsystem umfasst die Artikel und Hyperlinks zwischen den Artikeln. Die Artikel enthalten einen eindeutigen Titel und einen

Text, welcher die Bedeutung des enzyklopädischen Eintrags erklärt. Dabei gibt der erste Absatz eine kurze Definition des Begriffes wieder (Zech et al. 2007). Neben den üblichen Artikeln gibt es auch sogenannte „Begriffserklärungen“. Diese Artikel beschreiben mehrdeutige Begriffe und enthalten eine Liste aller Artikel, welche den Begriff mit seinen unterschiedlichen Bedeutungen beschreiben. Der Artikel „London (Begriffserklärung)“ umfasst eine Liste mit 37 Artikeln, die London (z.B. London, Hauptstadt von Grossbritannien oder London, eine amerikanische Heavy-Metal-Band) beschreiben.

Die Hyperlinks verbinden die Artikel untereinander und bilden ein gerichtetes Netzwerk. Eine Verbindung von einem Artikel zum anderen entsteht, wenn ein Begriff im Text eines anderen Artikels erwähnt wird. Diese Verbindungen drücken eine undefinierte semantische Beziehung zwischen den zwei Artikeln aus (Zech et al. 2007). Aus dem Artikel welcher „Zurich“ (Schweizer Stadt) beschreibt richten sich über 400 Links zu andern Artikeln, darunter z.B. „FIFA“, „*global city*“ und „*zurich german*“. Zürich ist der Sitz der FIFA, zählt nach gewissen Definitionen zu den globale Städten und die Hauptsprache ist *Züritüütsch*.

Den Artikeln werden aufgrund ihres Inhaltes, Schlagwörter, sogenannte Kategorien hinzugefügt. Die Schlagwörter klassieren die Artikel. Alle Artikel, die einer Kategorie zugewiesen werden, sind bezüglich dieses Aspekts ähnlich. So sind z.B. London, Berlin und Oslo mit dem Schlagwort „*capitals of Europe*“ versehen. Ein Artikel kann mit beliebig vielen Kategorien vermerkt werden. Die Verbindungen zwischen den Artikeln und den Kategorien sind die Verknüpfungen des Artikelsystems mit dem Kategoriensystem.

Das Kategoriensystem besteht aus Kategorien, welche in Form eines Thesaurus (Voss 2006) organisiert sind, d.h. dass zusätzlich zu den Schlagwörtern auch deren semantische Beziehungen explizit formuliert sind. Kategorien haben Verbindungen zu über- und untergeordneten Kategorien. Kategorien und Subkategorien stehen nach Zech et al. (2007) und Ponzetto und Strube (2007) in holonymischer Beziehung („hat-Beziehung“ z.B. Stadt > Quartier) und in hyperonymische Beziehung („ist-Beziehung“ z.B. Frucht > Pflaume) zu einander. Das Kategoriensystem ist halbstrukturiert (Ponzetto und Strube 2007). Konkret hat es keine genaue Baumstruktur, sondern Kategorien können mehreren Überkategorien untergeordnet sein und es gibt Schlaufen.

Semantische und geosemantische Analysen mit Wikipedia

Crowdsourcing-Projekte, wie Wikipedia und *OpenStreetMap* sind nicht nur für den alltäglichen Gebrauch von Interesse, sondern sind eine neue und herausfordernde Datenquelle für wissenschaftliche Analysen. Wikipedia, als mächtige Quelle für geographische Raum-Zeit-Untersuchungen (Goodchild 2007) und semantische Analysen (Zech et al. 2007) ermöglicht neues Wissen zu generieren und bestehendes Wissen zu validieren. Aus geoinformationswissenschaftlicher Perspektive spricht man in Bezug auf die freiwillige und kollektive Produktion von Daten mit geographischem Bezug von *volunteered geographic information* (VGI) (Goodchild 2007). Wikipedia wird bislang nicht als Geodatenbasis analysiert. Eine Ausnahme bilden die Arbeiten von Hecht (Hecht 2007; Hecht und Raubal 2008; Hecht und Moxley 2009; Hecht und Gergle 2010), die die Hyperlinks zwi-

schen den Artikeln im geosemantischen Kontext untersuchten mit dem Ziel Phänomene geographisch zu verorten. Semantische Ähnlichkeitsanalysen stützen sich sowohl auf den Inhalt der Artikel (Gabrilovich und Markovitch 2007), als auch auf das Kategoriensystem (Strube und Ponzetto 2006). Hecht (2007) stellt jedoch fest, dass durch die Analyse der strukturellen Elemente in Wikipedia (Hyperlink und Kategoriensystem) die Risiken (siehe oben) weniger ausgeprägt sind als bei reinen Textanalysen (Artikelinhalt).

Fazit

Wikipedia birgt ein riesiges Potential, nicht nur für linguistische und geoinformationswissenschaftliche sondern auch für soziogeographische Fragestellungen, wie sie diese Arbeit zu beantworten versucht. Dabei wird sich nicht nur die Aktualität des Wissens, sondern auch dessen soziale Verzerrung, die durch den *crowdsourcing*-Prozess entsteht, zu Nutze gemacht. Wikipedia ist eine Beschreibung unserer Welt aus der Perspektive der Menschen der Informationsgesellschaft, der *YOU*s nach Grossman (2006). Es ist ebenso ein Produkt der Informationsgesellschaft wie das globale Städtetz.

Die thematische Verzerrung und die mangelnde Wissenschaftlichkeit des Inhalts sind in dieser Perspektive nicht nur ein Risiko, sondern das Resultat sozialer Prozesse: Ereignisse, Personen und Themen, welche für die Informationsgesellschaft von Interesse sind, werden zueinander in Bezug gesetzt und bilden damit ein riesiges soziales Netzwerk. Wikipedia ist trotzdem keine demokratische und globale Darstellung unserer Gesellschaft, weil sich das *YOU* zwar sehr umfassend anhört, sich in der Realität jedoch insbesondere auf junge, gebildete Männer ohne Kinder (Glott et al. 2010) beschränkt.

Die Risiken der mangelnden Genauigkeit und der Volatilität müssen berücksichtigt werden. Eine hohe Autorenzahl wirkt sich positiv auf die Genauigkeit aus, erhöht jedoch gleichzeitig die Volatilität der Beiträge. Durch Sensitivitätsanalysen können die Risiken nicht wettgemacht, aber zumindest quantifiziert werden. Da der Raum in der gestellten Fragestellung von zentraler Bedeutung ist, sind die sprachlichen und insbesondere die räumlichen Verzerrungen auch mittels Sensitivitätsanalysen zu ermitteln.

Unter Berücksichtigung der Feststellung von Hecht (2007), dass strukturelle Elemente vorzuziehen sind, fokussiert sich das Interesse auf die Identifikation eines funktionalen differenzierten Städtetzes mittels Hyperlinks und Kategoriensystem. Artikel- und Kategoriensystem sind in den monatlich gespeicherten Versionen von Wikipedia (*XML-Format*) enthalten. Die Version der englischen Wikipedia vom 13. Januar 2010 dient als Datengrundlage. Die englische Version wird unter anderem wegen ihrer Grösse den anderen vorgezogen. Mit dem von Hecht (2007) in *Java* entwickelten *WikiAPI* wird das *XML-File* in eine MySQL-Datenbank in strukturierter Form geschrieben (Tabelle 14).

Tabelle 14: Aus der englischen Wikipedia extrahierte und in der Arbeit verwendete Daten

Tabelle	Beschreibung	Variablen	Anzahl Einträge
Artikelsystem			
local_articles	Liste der Artikel	Local_id, title, inlinks, outlinks	3'206'965
links	Liste der Verbindungen zwischen den Artikeln	start, end, start_title, end_title	94'711'049
ills	Liste der Verbindungen zwischen englischen Artikeln und gleichen Artikeln in anderen Sprachen.	source_title, source_lang, dest_title, dest_lang	7'294'959
Kategoriensystem			
local_categories	Liste der Kategorien	local_id, title	551'540
category_links	Liste der Verbindungen zwischen den Kategorien	child, parent	1'059'166
category_memberships	Liste der Verbindungen zwischen Artikeln und Kategorien	Local_article_id, local_category_id	10'516'272

5 Methodik

Das methodische Vorgehen ist in vier Hauptbereiche unterteilt und stellt die Konkretisierung des methodischen Rahmens für die Untersuchung und Analyse des globalen Städtenetzes mittels Wikipedia dar. Die Städte und deren Relationen sind Gegenstand des ersten Teils. Die Definition des Städtenetzes in Wikipedia, seine Modellierung und die visuell-analytische Exploration desselben stehen dabei im Vordergrund. Der zweite Schritt befasst sich mit den Akteuren und deren Eigenschaften. Die Akteure werden aus Wikipedia extrahiert und in Hinsicht auf ihre Funktion analysiert. Im dritten Bereich werden die vorgängigen Ergebnisse kombiniert und das funktional differenzierte Städtenetz analysiert. Im letzten Teil wird die Methodik zur Evaluation von besonders relevanten Arbeitsschritten und der Datengrundlage vorgestellt.

5.1 Modellierung des allgemeinen Städtenetzes

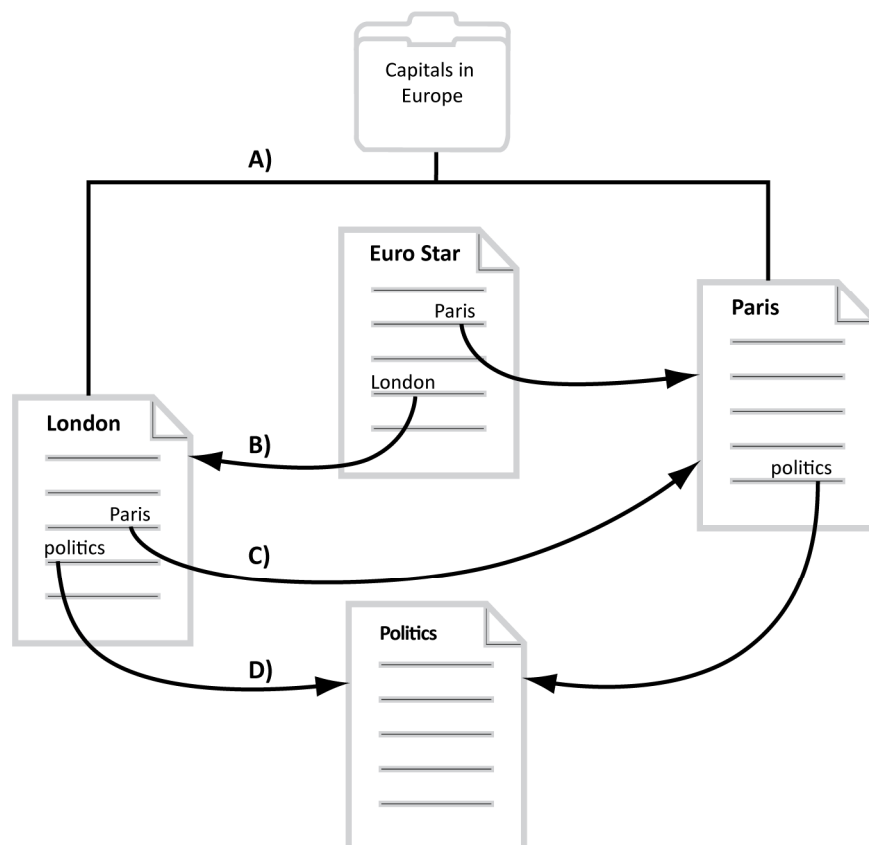
Städte gehören zum enzyklopädischen Wissen und sind somit als Artikel in Wikipedia vorhanden und eindeutig identifizierbar. Sie werden aufgrund des Artikels „*List of towns and cities with 100'000 or more inhabitants*“⁵ bestimmt. Die von diesem Artikel ausgehenden Links sind mehrheitlich auf Artikel mit Städten gerichtet, die potentiell globale Städte sind. Die Bevölkerungszahl ist kein entscheidendes Kriterium für die Definition einer globalen Stadt. Da 100'000 eine verhältnismässig geringe Bevölkerungszahl ist, kann man davon ausgehen, dass alle globalen Städte dieses Mass übertreffen und sich somit

⁵ Als Quelle für die Liste der Städte gilt *United Nations Statistics Division* (2006).

dieses Kriterium für eine erste Abgrenzung des Systems eignet. Aus der *local_articles*-Tabelle werden 3'532 Artikel extrahiert.

Die Relationen zwischen diesen Städten bzw. Artikeln können auf unterschiedliche Weise definiert werden (Abbildung 12). Alle vier Varianten in diesem Beispiel stellen eine Beziehung zwischen London und Paris dar. Allerdings lässt sich nur ausgehend von den einkommenden Hyperlinks von einem gemeinsamen Artikel (Variante B) eine funktional differenzierte Beziehung zwischen zwei Städten definieren. Nach Hecht (2007) sind Beziehungen über Hyperlinks inhaltlich vielfältiger und geeigneter für die Untersuchung von räumlichen Prozessen, als Beziehungen über Kategorien. Über Kategorien (Variante A) entstehen klassische hyperonymische/hyponymische und meronymische/holonymische Relationen, z.B. London und Paris sind verbunden, weil beide Hauptstädte Europas sind. Über dem Hyperlinknetzwerk entstehen hingegen auch nicht-klassische Beziehungen, z.B. zwischen Paris und London fährt der Euro Star oder Paris und London sind von grosser politischer Relevanz.

Abbildung 12: Mögliche Relationen zwischen zwei Artikeln in Wikipedia. A) Mit gleicher Kategorie markiert. B) Einkommender Hyperlink aus gemeinsamem Artikel. C) Direkter Hyperlink von einem Artikel zum anderen. D) Ausgehender Hyperlink zu gemeinsamem Artikel

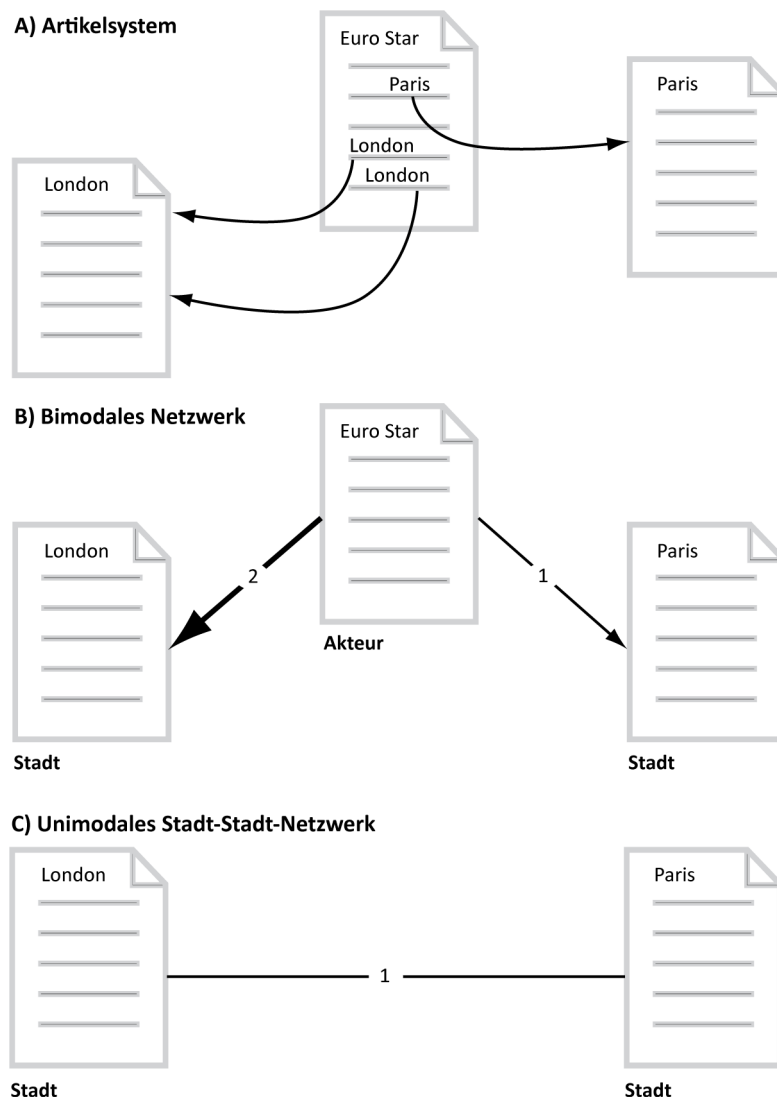


Die Vorteile von Variante B) gegenüber den weiteren Beziehungen über Hyperlinks sind sowohl datentechnischer als auch konzeptioneller Natur. Variante B) ist robuster: die Hyperlinks haben ihren Ursprung in ganz unterschiedlichen und meistens ganz vielen Artikeln, die von unterschiedlichen AutorInnen geschrieben worden sind. Diese höhere

Menge und Vielfalt reduzieren die Risiken und erhöhen die Zuverlässigkeit und Robustheit der Resultate. Aus konzeptioneller Sicht entstehen die Städte durch das Handeln von Akteuren, welche Flüsse von Menschen, Informationen und Güter generieren. Ausgangspunkt sind somit die urbanen Akteure und nicht die Städte an sich. Dieses Sachverhältnis wird von Variante B am Besten wiedergegeben; die verbindenden Artikel sind aus dieser Perspektive urbane Akteure.

Das resultierende Netzwerk ist bimodal und gerichtet (Abbildung 13). Die einen Knoten umfassen die Städte, die anderen die urbanen Akteure. Die Kanten verlaufen stets von einem Akteur zu einer Stadt und haben unterschiedliche Gewichte (B), die der Anzahl Links von einem Akteur-Artikel zu einem Städte-Artikel entsprechen. In diesem bimodalen Netzwerk ist latent ein unimodales Stadt-Stadt-Netzwerk (C) vorhanden (Abbildung 13).

Abbildung 13: Umwandlung des Artikelsystems in Wikipedia zu einem unimodalen Stadt-Stadt-Netzwerk



Die Umwandlung des bimodalen in ein unimodales Netzwerk basiert auf dem *max-flow-min-cut*-Theorem (Ford und Fulkerson 1956). Das Theorem besagt, dass der maximale

Fluss zwischen zwei Knoten dem minimalen Schnitt entspricht, welcher die zwei Knoten vollkommen trennt. Der minimale Schnitt entspricht den Flaschenhälsen entlang eines Pfades. Bei der vorliegenden Transformation ist die relevante Pfadlänge zwischen zwei Städten immer zwei. Der minimale Schnitt bzw. das Kantengewicht im unimodalen Netzwerk entspricht dem kleinsten Gewicht im Kantenpaar des Pfades (Formel 1).

Formel 1: Transformation des bimodalen Netzwerks in ein Stadt-Stadt-Netzwerk mit Anwendung des *max-flow-min-cut*-Theorems

$$KG_{ij} = \sum_{r=1}^n \min(KG_{ri}, KG_{rj})$$

wo KG_{ij} das Kantengewicht zwischen Stadt i und Stadt j im Stadt-Stadt-Netzwerk ist. KG_{ri} stellt das Kantengewicht zwischen Stadt i und dem Akteur r und KG_{rj} das Kantengewicht zwischen demselben Akteur und der Stadt j im bimodalen Netzwerk dar.

Die Akteur-Artikel haben sehr unterschiedliche Längen und damit eine variierende Anzahl ausgehender Links. Um Verzerrungen zu vermeiden drängt sich eine Normalisierung aufgrund der Anzahl ausgehender Links aus dem Akteur-Artikel auf. Hecht und Raubal (2008) entwickeln ein Mass, welches die semantische Distanz zwischen zwei Artikeln in Wikipedia wiedergibt. Eine leicht modifizierte Formel kann für die Berechnung der Kantengewichte angewendet werden (Formel 2).

Formel 2: Berechnung des normalisierten Kantengewichtes in Anlehnung an Hecht und Raubal (2008)

$$\text{Für } C \leq 5 \quad NKG_{ij} = \frac{KG_{ij}}{OD_r}$$

$$\text{Für } C > 5 \quad NKG_{ij} = \frac{KG_{ij}}{C + (1 + \log_2 |OD_r - C|)}$$

wo NKG das normalisierte Kantengewicht zwischen den Städten i und j , KG das nicht-normalisierte Kantengewicht zwischen denselben, OD das *outdegree* des Akteurs r sind. C entspricht einer Konstante welche der Anzahl ausgehender Links entspricht.

Das Logarithmieren des *outdegrees* und der Einbezug der Konstante stützen sich auf vertiefte Erfahrungen mit Wikipedia. Damit wird berücksichtigt, dass lange Artikel in thematische Absätze unterteilt und vielfältiger sind (Hecht und Raubal 2008). Ein gutes Beispiel dafür sind Artikel über Länder, die sowohl Absätze zu Klima und Natur, als auch zu Politik und Wirtschaft enthalten. Die Konstante, welche erfahrungsgemäss gleich fünf gesetzt wird, entspricht der erwarteten Anzahl Links aus einem thematischen Absatz. Der Logarithmus approximiert die Links in allen anderen Absätzen (Hecht und Raubal 2008).

Die Verbindung zwischen zwei Städten besteht aus der Summe der normalisierten Kantengewichte über alle Akteure, die diese zwei Städte verbinden. Das Resultat der Modellierung ist eine gewichtete ungerichtete Adjazenzmatrix mit 3'532 Knoten und maximal

6'235'746 Kanten. Diese Matrix ist die Grundlage für die Berechnung aller Merkmale des Netzwerks.

5.2 Analyse des allgemeinen Städtetetzes

Zentralität

Die Zentralitätsmasse sind die relevantesten Kennwerte auf individueller Ebene und beschreiben die Wichtigkeit eines Knotens aus relationaler Perspektive. Bei ungerichteten Graphen ist die Zentralität eine Funktion der Einbindung bzw. der Beteiligung eines Knotens an vielen Beziehungen (Wasserman und Faust 2008). Diese Masse finden in der Analyse des globalen Netzwerkes breite Anwendung. Ein Beispiel dafür sind Alderson und Beckfield (2004).

Die *degree* Zentralität ist das einfachste Mass um die Einbindung eines Knotens zu ermitteln. Freeman (1979) definiert den *degree* als Anzahl Beziehungen eines Knotens bzw. die Summe der Kantengewichte eines Knotens für gewichtete Netzwerke (Formel 3). Der *degree* berücksichtigt ausschliesslich die direkten Beziehungen und gibt die Aktivität und die lokale Einbettung eines Knotens wieder. Je höher der *degree*, desto aktiver bzw. desto verknüpfter ist der Knoten im Netzwerk. (Wasserman und Faust 2008)

Formel 3: Gewichteter *degree* für einen Knoten (n_i).

$$C_D(n_i) = \sum_j w_{ij}$$

wo w_{ij} die Kantengewichte zwischen dem Knoten i und allen adjazenten Knoten j sind.

Die nähebasierte Zentralität (*closeness centrality*) (Freeman 1979) erfasst nicht das Volumen der direkten oder indirekten Verbindungen, sondern deren Länge (Borgatti und Everett 2006). Die nähebasierte Zentralität ist als die Summe der Länge der geodesischen Pfade von einem Knoten zu jedem anderen im Netzwerk definiert. Die dahinterstehende Idee ist, dass ein zentraler Knoten schnell jeden anderen erreicht und mit ihm interagieren kann (Wasserman und Faust 2008). Im vorliegenden Städtetetz sind in der Tendenz alle Knoten direkt mit allen anderen verbunden; die ungewichtete nähebasierte Zentralität wäre somit für alle Knoten gleich. Die Berücksichtigung aller Pfade zwischen zwei Knoten ergibt einen globaleren und umfassenderen Einblick in die Erreichbarkeit der Knoten. Mit der Anwendung vom *max-flow-min-cut*-Theorem (Algorithmus von Gomory und Hu 1964: in UNICET 6) lässt sich der maximale Fluss zwischen zwei Knoten ermitteln. Dieser entspricht in gewichteten Netzwerken der Summe der Kantengewichte der schwächsten Kanten in jedem der möglichen Pfade zwischen zwei Knoten (Ford und Fulkerson 1956). Der maximale Fluss kann als eine Distanzmetrik unter Berücksichtigung aller Pfade interpretiert werden. Die nähebasierte Zentralität eines Knotens ist somit die Summe der maximalen Flüsse, die vom betrachteten Knoten ausgehen.

Der *degree* und die nähebasierte Zentralität sind radiale Zentralitätsmasse, d.h. sie geben die Kohäsion eines Knotens mit den anderen Knoten im Netzwerk und die Stärke des

Einflusses eines Knotens auf die Netzwerkstruktur wieder. Radiale unterscheiden sich von medialen Massen. Letztere betrachten nicht Pfade, die von einem Knoten ausgehen, sondern Pfade, die durch einen Knoten hindurch verlaufen. Die *betweenness* Zentralität ist ein solches mediales Mass. Die *betweenness* Zentralität drückt die Möglichkeit eines Knotens, die Interaktionen zwischen anderen Knoten zu kontrollieren, aus. (Wasserman und Faust 2008)

Die *betweenness* Zentralität wird in gewichteter Form mit dem Algorithmus nach Brandes (2001b), welcher in Tulip 3.5 (LaBRI 2010) implementiert ist und auch für sehr grosse und gewichtete Netzwerke taugt, berechnet.

In der Regel ist die Korrelation zwischen den unterschiedlichen Zentralitätsmassen, trotz unterschiedlicher Annahmen und Berechnungen, hoch. Der Vergleich dieser Werte ergibt jedoch, gerade wenn diese Regularität nicht gegeben ist, sogenannte Anomalien, interessante Hinweise auf die Struktur des Netzwerks und die Eigenschaften der einzelnen Knoten. Guimerà et al. (2005) vergleichen den *degree* mit der *betweenness* Zentralität und identifizieren eine *degree-betweenness*-Anomalie: Einzelne wenige Knoten haben eine hohe *betweenness* Zentralität, trotz einem eher tiefen *degree*. Diese Anomalie ist ein Zeichen für Subgruppenstrukturen im Netzwerk. Die anomalen Knoten sind dabei verbindende Knoten zwischen den Subgruppen, sie erfüllen eine Brückenfunktion und spielen eine zentrale Rolle beim Zusammenhalten von Netzwerken und bei Diffusionsprozessen. (Guimerà et al. 2005)

Die Struktur des Netzwerks lässt sich auch mittels der Analyse der Zentralitätsverteilung mit einem Log-Log-Graph eruieren (Pumain 2006a): die y-Achse gibt den Logarithmus der Zentralitätswerte der einzelnen Knoten und die x-Achse den Logarithmus der Rangfolge wieder. Die Steigung der Kurven zeigt die hierarchische Differenzierung: je grösser die Steigung desto stärker die Differenzierung.

Kohäsive Subgruppen und Rollen

Kohäsive Subgruppen und Rollen sind in erster Linie Elemente der Gruppenebene, können jedoch auch als Eigenschaften der einzelnen Knoten betrachtet werden. Für die erste Perspektive liegt der Fokus auf den Subgruppen bzw. Rollen und deren Eigenschaften. Für die zweite hingegen sind die Rollen und Subgruppenzugehörigkeiten Attribute der Knoten.

Es existiert eine Vielzahl von Definitionen von Subgruppen. Diese lassen sich auf vier Konzeptionen zurückführen. Eine erste definiert Subgruppen aufgrund der Gegenseitigkeit: alle Elemente einer Subgruppe sind gegenseitig verbunden. Schwächer sind die Bedingungen für Subgruppen, welche auf die Erreichbarkeit zwischen den Knoten einer Subgruppe basieren. Knoten müssen nicht adjazent sein, eine bestimmte maximale Pfadlänge zu den anderen Gruppenmitgliedern ist ausreichend. Weitere Definitionen stützen sich auf eine minimale Anzahl von direkten Verbindungen, die jeder Knoten zu anderen Mitgliedern derselben Subgruppe aufweisen muss. Die mathematisch kom-

plexesten Definitionen beziehen sich auf den Vergleich der Beziehungen innerhalb und zwischen den Gruppen. (Wasserman und Faust 2008)

Der *blondel community detection* Algorithmus (Blondel et al. 2008) stützt sich auf diese letzte Konzeption. Die Identifizierung der Subgruppen basiert auf der Modularitätsoptimierung. Die Modularität drückt das Verhältnis zwischen der Dichte der intra- und inter-Subgruppenbeziehungen aus (Sales-Pardo et al. 2007) und ist nach Newman (2004) auch auf gewichtete Netzwerke anwendbar (Formel 4). Der Blondel-Algorithmus ist für die Erkennung von Subgruppen in grossen, gewichteten und realen Netzwerken entwickelt worden und in NWB (NWB Team 2006) verfügbar. Die Anzahl Subgruppen ist nicht a priori definiert und jeder Knoten gehört zu genau einer Subgruppe. Durch mehrere Durchläufe werden einzelne Subgruppen zu übergeordneten Subgruppen zusammengefasst, womit die hierarchische Struktur des Netzwerkes ermittelt wird. (Blondel et al. 2008)

Formel 4: Modularität für einzelne Subgruppen nach Newman und Girvan (2004) angepasst für gewichtete Netzwerke und normalisiert

$$M_i = \frac{\frac{l_i}{L} - \left(\frac{d_i}{2L}\right)^2}{N_i}$$

wo l_i die Summe der Intragruppenkantengewichte in Gruppe i , L die Summe der Kantengewichte im Netzwerk, d_i die Summe der gewichteten *degree* aller Knoten in Gruppe i und N_i die Anzahl Knoten in i sind.

Rollen geben einen Einblick in die innere Struktur der Subgruppen und indirekt eine umfassende Beschreibung der Netzwerkstruktur. Sie setzen keine direkte Verbindung zwischen den Knoten voraus, sondern werden aufgrund ähnlicher Beziehungsmuster gebildet (Wasserman und Faust 2008). Guimerà und Amaral (2005a) definieren eine kleine Anzahl universelle, für alle Netzwerke geltende Rollen. Die Rollen basieren auf Modulen im Netzwerk. Diese entsprechen den kohäsiven Subgruppen (Guimerà und Amaral 2005b).

Für die Berechnung der Rollen im Städtetz wird die höchste hierarchische Ebene der oben mit dem Blondel-Algorithmus definierten Subgruppen berücksichtigt. Nach Guimerà und Amaral (2005b) werden die Rollen vom intramodularen *degree* und dem Partizipationskoeffizienten definiert. Diese Masse werden an das gewichtete Netzwerk angepasst. Dabei werden die Kantengewichte mitberücksichtigt, weil diese im Städtetz besonders informationsreich sind. Der intramodulare gewichtete *degree*, Z-Wert genannt, wird nach Formel 5a berechnet und besagt, wie gut ein Knoten mit den anderen in derselben Subgruppe verbunden ist. Der Partizipationskoeffizient (P-Wert, Formel 5b) gibt die Beziehungen eines Knotens zu den fremden Subgruppen wieder. Der P-Wert ist nahe bei null, wenn der Knoten wenig Kanten zu Subgruppen-externen Knoten hat. (Guimerà und Amaral 2005b)

Formel 5: Z-Wert A) und P-Wert B) für die Bestimmung der Rollen (Guimerà und Amaral 2005b)

$$Z_i = \frac{k_i - \bar{k}_{si}}{\sigma_{k_{si}}} \quad \text{A)}$$

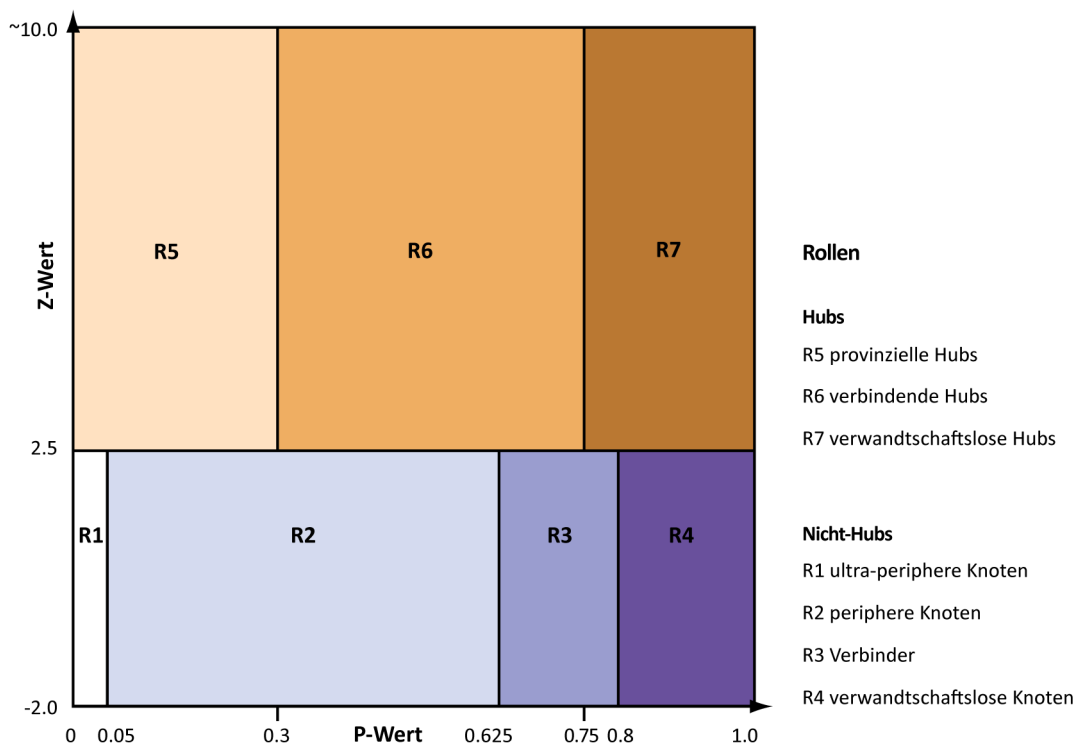
wo k_i die Summe der Kantengewichte von Knoten i zu allen Knoten in derselben Subgruppe s_i ist. \bar{k}_{si} ist der Durchschnitt und $\sigma_{k_{si}}$ die Standardabweichung von k über alle Knoten in s_i .

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{k_{is}}{k_i} \right)^2 \quad \text{B)}$$

wo k_{is} die gewichtete Summe der Kanten von Knoten i zu den Knoten in alle (von 1 bis N_M) Subgruppen s und k_i der gewichtete *degree* von i sind.

Aufgrund des Z-Wertes und des P-Wertes werden die Knoten in sieben unterschiedliche Rollen unterteilt. Der P-Wert unterscheidet periphere (tiefer P-Wert) von verbindenden Knoten (hoher P-Wert) und der Z-Wert unterteilt die Knoten in Hubs (hoher Z-Wert) und Nicht-Hubs (tiefer Z-Wert) (Abbildung 14).

Abbildung 14: Rollendefinition aufgrund von Z-Werten und P-Werten (Guimerà und Amaral 2005b)



Die innere Struktur der Gruppen lässt sich auch mittels des Homophilie-Indexes explorieren. Dabei wird für jeden Hub der Anteil an Intragruppenbeziehungen, gemessen am gewichteten *degree*, berechnet. Je höher der Anteil ist, desto höher die Homophilie und desto stärker die innere Ausrichtung einer Stadt ist. Ergänzend dazu wird die Verteilung der Beziehungen eines Knotens pro Gruppe eruiert. Dies zeigt, zu welchen anderen Gruppen, abgesehen von der eigenen, eine Stadt besonders starke bzw. besonders schwache Beziehungen pflegt.

Layout

Die Visualisierung des Stdtenetzes richtet sich nach der in Kapitel 3.2 entwickelten Netzwerkvisualisierung. Dabei werden zwei unterschiedliche Geometrien erzeugt. Der erste topologische Raum wird aufgrund einer Inputmatrix mit den 3'532 Stdten und deren 15 strksten Beziehungen berechnet. Die Dimensionsreduktion sttzt sich auf den DrL/VxOrd-Algorithmus mit einem *edge cutting*-Wert von 0.8. Die Punkt-Visualisierung wird geometrisch generalisiert und die Trajektorien entsprechen dem *pathfinder network* mit $r=\infty$ (*fast pathfinder network scaling* in NWB) gefiltert. Die ursprngliche Adjazenzmatrix (3'532 Knoten und 2'506'284 Kanten) wird in 63 Aggregate und 62 Trajektorien zusammengefasst.

In einem zweiten Layout liegt der Fokus auf den Hubs im globalen Stdtenetz. Fr das Projizieren des zweiten Layouts werden ausnahmslos Hubs nach Guimer und Amaral (2005a) (Rollen R5-R7) ausgewhlt. Die Inputmatrix besteht ausschliesslich aus dem Pathfinder Network mit $r=\infty$ zwischen den Hubs. Die Geometrie des resultierenden Netzwerkes, welches die Hubs und die Hauptachsen des globalen Stdtesystems darstellt, wird mit dem ungewichteten GEM-Algorithmus generiert. Die Punkt-Visualisierung wird nicht geometrisch generalisiert. Die ursprngliche Adjazenzmatrix (3'532 Knoten und 2'506'284 Kanten) wird in 95 Loci (Hubs) und 94 Trajektorien zusammengefasst.

Die Kantengewichte werden anhand der Farb-Muster-Variablen Grsse und Helligkeit (Fabrikant et al. 2004) dargestellt. Starke Verbindungen sind mit breiten, dunklen Linien und schwache Verbindungen mit schmalen, hellen Linien dargestellt.

Die Zentralittswerte, Rollen und Subgruppenzugehrigkeit sind alles Attribute der Knoten und werden somit mit der Variation der graphischen Variablen der Punkte abgebildet: Zentralitt mit der Grsse, Rolle mit der Form und Subgruppen mit dem Farbton.

5.3 Differenzierung der Funktionen

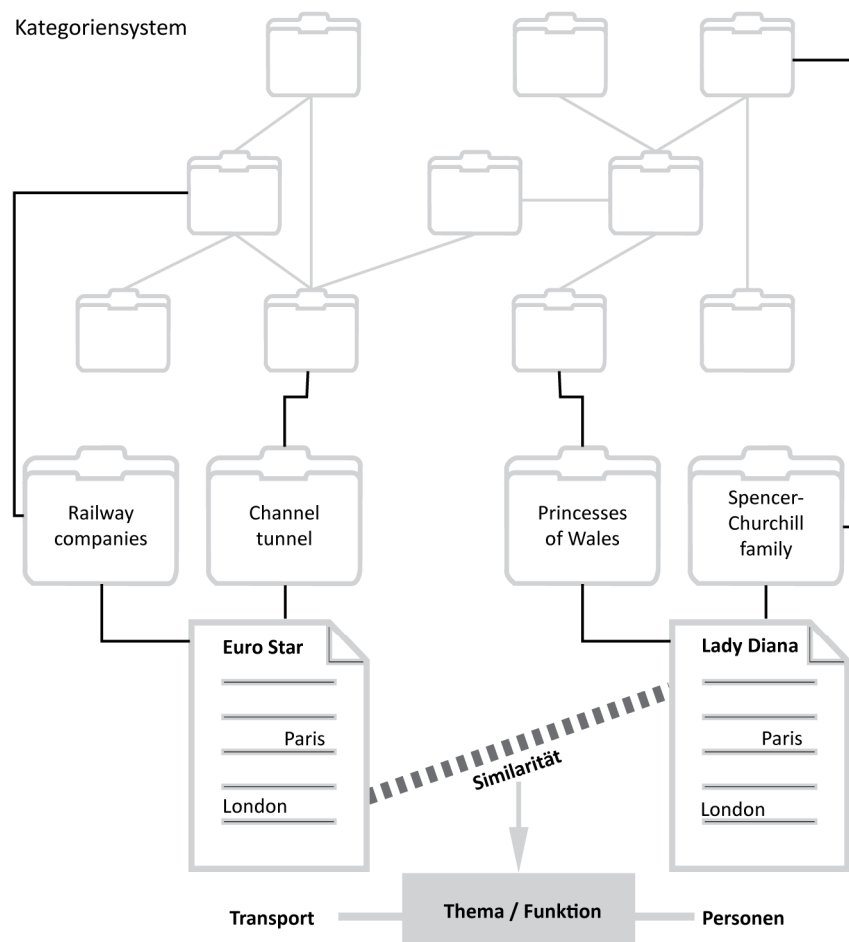
Der zweite Schritt befasst sich nicht mehr mit den Stdten an sich, sondern mit den Akteuren, welche die Relationen zwischen den Stdten aufspannen. Ziel ist es, die Funktion dieser Akteure zu identifizieren. Aufgrund der grossen Datenmenge wird nur ein Teilbereich des Netzes ausgewhlt: die folgenden Analysen konzentrieren sich auf das Netz unter den 95 Hubs. Dieses Netz wird aufgrund von 225'701 Artikeln aufgespannt, fr die es eine Funktion zu definieren gilt.

Diese Artikel sind mit Kategorien markiert, die den Inhalt schlagwortartig beschreiben. Die Markierung der Artikel mittels Kategorien basiert mehrheitlich auf hyperonymischen/hyponymischen Beziehungen und ermglicht das Herleiten von semantischen Similarittsbeziehungen (Abbildung 15) (Holloway et al. 2007). Die semantische hnlichkeit ist eine Untermenge der semantischen Verbundenheit (Budanitsky und Hirst 2006). Semantische Verbundenheit umfasst weitere klassische und nicht-klassische Ty-

pen von semantischen Beziehungen (z.B. Synonyme, Meronyme) (Budanitsky und Hirst 2006; Hecht und Raubal 2008) und ist über die Verlinkung zwischen den Artikeln gegeben. Für die Differenzierung der Funktionen ist die semantische Ähnlichkeit entscheidend.

Die semantische Ähnlichkeit lässt sich mit dem direkten Vergleich der Kategorien ermitteln. Dadurch sind Aussagen wie „zwischen den Artikeln Lady Diana und Euro Star gibt es keine Übereinstimmung der Kategorien“ möglich. Durch den Einbezug des gesamten Kategoriensystems sind nicht nur binäre Aussagen möglich (zwei Kategorien sind gleich oder ungleich), sondern differenziertere Ähnlichkeiten sind definierbar. Konkret könnte die Ähnlichkeit zwischen zwei Kategorien, und somit indirekt zwischen zwei Artikeln, aufgrund der Pfaddistanz zwischen denselben im Kategoriensystem definiert werden: je näher zwei Kategorien sind, desto ähnlicher sind sie (Abbildung 15).

Abbildung 15: Typisierung der verbindenden Artikel aufgrund der Markierung mit Kategorien



Die Ermittlung der Ähnlichkeit stützt sich auf die drei Tabellen des Kategoriensystems (*local_categories*, *category_links*, *category_memberships*) und folgt dem Prinzip, welches auch bei semantischen Analysen, wie dem *topic modelling* (Steyvers und Griffiths 2007) gilt. Die Artikel werden einer bestimmten Anzahl Themen gegenüber gestellt und aufgrund der Zugehörigkeit der Artikel zu den unterschiedlichen Themen wird deren Ähnlichkeit abgeleitet.

Die Tabelle *category_links* enthält in der Form einer halbhierarchischen Struktur die Relationen zwischen den Kategorien und wird zur Definition der Themen genutzt. Zwei der obersten Kategorien dieser Hierarchie kommen einer Themendefinition gleich: „*fundamental categories*“ (Wikipedia contributors 2011b) und „*main topic classifications*“ (Wikipedia contributors 2011a). Die erste umfasst die vier grundsätzlichen ontologischen Entitäten: soziale, biologische, physische und intellektuelle Entitäten. Die zweite hingegen teilt sich in ca. 20 Hauptthemen, die breit genug gewählt sind, dass jede einzelne Kategorie über das Kategoriensystem mindestens zu einem dieser Themen gelangt (Wikipedia contributors 2011a). Die Themen in der „*main topic classifications*“ entsprechen den thematischen Schwerpunkten, welche im Städtenetz erwartet werden, besser. Bezogen auf die „*fundamental categories*“ würden die meisten Funktionen im Städtenetz als soziale Entitäten klassiert werden.

Ausgangspunkt der Themendefinition sind somit die Subkategorien der Hauptkategorie „*main topic classifications*“ (Tabelle 15). Aus der Erfahrung mit dem *topic modelling* ist bekannt, dass eine höhere Anzahl Themen die semantische Struktur der Daten besser eruiert. Aus diesem Grund werden die Themen um die Subkategorien dieser 22 Hauptthemenkategorien erweitert. Jedoch werden nur Subkategorien, welche mindestens selber 20 Subkategorien haben und aus einem bzw. drei Wörtern bestehen, wobei das mittlere ein „*and*“ ist, als eigenständige Themen betrachtet. Mit dem Ausschliessen von Kategorien mit wenigen Subkategorien wird vermieden, dass zu spezifische Themen definiert werden. Bei der Ausscheidung von Kategorien, welche aus mehreren Begriffen bestehen, folgt man dem Grundsatz, dass man ein wichtiges Thema mit einem gegebenen Begriff definieren kann. Kategorien, wie „*people by nationality*“ oder „*social science*“ werden somit nicht als eigenständige Themen definiert, sondern zählen zu „*people*“ bzw. „*science*“. Als Endresultat ergeben sich 83 Themen.

Tabelle 15: Hauptthemen in Wikipedia definiert anhand der Kategorie „*Main topic classifications*“

Themen	Anzahl Subkategorien	Themen	Anzahl Subkategorien
agriculture*	55	history*	40
arts*	25	humanities*	18
belief*	11	language*	21
business*	57	law*	46
chronology*	17	mathematics*	55
computing*	39	nature*	20
culture*	51	people*	25
education*	52	politics*	29
environment*	32	science*	39
geography*	24	society*	44

Themen	Anzahl Subkategorien	Themen	Anzahl Subkategorien
health*	33	technology*	51
academia	32	industry	30
aesthetics	20	internet	23
analysis	22	labor	22
anthropology	33	languages	33
archaeology	32	linguistics	49
architecture	37	literature	54
buildings_and_structures	23	logic	23
businesspeople	20	management	35
centuries	73	marketing	41
clothing	20	measurement	28
clubs_and_societies	50	military	27
comedy	25	months	20
commerce	23	optics	30
communication	42	organizations	21
companies	25	philosophy	23
crafts	28	places	26
crime	26	psychology	29
decades	399	religion	42
disability	24	scientists	61
drama	23	sociology	29
economics	34	software	40
elections	36	spirituality	21
engineering	26	sports	40
entertainment	47	sustainability	36
equipment	26	systems	20
events	20	timelines	24
forestry	20	topology	22
geometry	38	traditions	25
government	46	writing	24
healthcare	33	years	2755
historiography	22	* Direkte Subkategorien von "Main topic classifications"	

Als nächster Schritt werden alle Kategorien im Kategoriensystem diesen Themen zugewiesen und zwar mit Hilfe der fast hierarchischen Beziehungen zwischen denselben. Jede Kategorie wird demjenigen Thema zugewiesen, zudem sie den kürzesten Pfad hat. Wenn der Pfad zu mehreren Themen gleich lang ist, gehört die Kategorie zu mehreren Themen. Für jede Kategorie wird auch ein Gewicht berechnet, welches die Spezifität der Zugehörigkeit wiedergibt: gehört eine Kategorie ausschliesslich zu einem Thema ist das Gewicht eins. Um das Gewicht zu berechnen, werden auch multiple Pfade von einer Kategorie zum selben Thema berücksichtigt: ist eine bestimmte Kategorie mit dem Thema „*arts*“ über vier gleichlange Pfade und mit dem Thema „*technology*“ über zwei gleichlange Pfade erreichbar, dann ist das Gewicht gleich zwei Drittel für „*art*“ und ein Drittel für „*technology*“. Bei den Analysen hat sich herauskristallisiert, dass eine Lockerung der Bedingung der gleichen Pfadlänge die Mehrdeutigkeit der Kategorien besser widerspiegelt. Der Grund dafür ist die heterogene Struktur des Kategoriensystems. Dies lässt sich gut an einem Beispiel erläutern. Die zu definierende Kategorie „*FC Roma*“ ist der Kategorie „*geography by country*“ unterstellt, welche der Kategorie „*geography*“ unterstellt ist. Andererseits ist die Kategorie „*FC Roma*“ der Kategorie „*sport by city*“ unterstellt, welche wiederum der Kategorie „*sport by country*“ und sie desgleichen der Kategorie „*sport*“ unterstellt ist. Die unterschiedlichen Pfadlängen zwischen „*FC Roma*“ und „*sport*“ einerseits und „*geography*“ andererseits, sind die Konsequenz unterschiedlicher Detaillierungsgrade. Daraus lässt sich schliessen, dass kleine Unterschiede in der Pfadlänge noch kein Indiz auf unterschiedliche semantische Ähnlichkeit sind. Aus diesem Grund müssen die Pfadlängen nicht exakt gleich sein, sondern können um einen Schritt variieren.

Die Verknüpfung der Artikel zu den 83 Themen erfolgt über die MySQL-Tabelle *category_memberships*, in der die Beziehungen zwischen jedem Artikel und den Kategorien enthalten sind. Die Kategorien beschreiben in Form von Tags den Inhalt der Artikel (Voss 2006). Nicht alle Kategorien werden berücksichtigt: Kategorien mit ausschliesslich zeitlichen oder räumlichen Inhalten werden nicht zur thematischen Definition der Artikel verwendet. Konkret werden Millenniums-, Jahrhundert-, Jahres- (z.B. „2002“), Monats- und Tageskategorien (z.B. „9. April“), sowie Kontinent-, Länder-, Regionen- und Städte-kategorien gelöscht. Damit wird die Gruppierung der Artikel um raum-zeitliche Eigenschaften reduziert. Die hier diskutierte Klassifizierung legt das Schwergewicht auf thematisch-funktionale Aspekte.

Die 234'059 Artikel werden mit den 83 Themen verbunden. Wegen fehlender Links im Kategoriensystem, gehören 8'358 Artikel (3.6%) zu keinem Thema. Die vorhin definierten Gewichte zwischen Kategorien und Themen werden durch die Anzahl Kategorien, zu denen ein Artikel gehört, normalisiert. Die Summe der Gewichte zwischen einem Artikel und den Themen ist immer gleich eins.

Das resultierende Netzwerk ist ein bimodales Artikel-Themen-Netzwerk. Von den unterschiedlichen analytischen Ansätzen wird eine Kombination zwischen SOM und Netzwerkanalyse gewählt. Das Netzwerk ist durch eine sehr hohe Dichte gekennzeichnet, was eine Netzwerkvisualisierung beträchtlich erschwert. Die Filterung nach Klavans

und Boyack (2006a), nach der nur die wichtigsten Beziehungen pro Knoten berücksichtigt werden und die folgende Projektion im Raum mit dem DrL/VxOrd-Algorithmus (Shawn et al. 2011), vermögen die Strukturen des bimodalen Netzwerks nicht visuell hervorzuheben. Die Visualisierung mittels SOM ist eine optimale Alternative. Nicht nur aus datentechnischen Gründen ist die Methode erfolgversprechend, sondern sie gibt die Inhalte auch metaphorisch optimal wieder. Die Container-Metapher hebt die Zugehörigkeit eines Artikels zu einem Thema besonders hervor. Mit dem SOM Analyst (Lacayo-Emeri 2011b) werden die 234'059 Artikel aufgrund ihrer Beziehung zu den Themen in eine SOM mit 100 x 100 Neuronen projiziert. Die Projektion in der SOM mit 100 x 100 Neuronen führt sowohl zu einer Dimensionsreduktion (83 Themen zu 2 Dimensionen) als auch zu einer Klassierung (234'059 Artikel zu 10'000 Neuronen). Dadurch wird die Lesbarkeit der Daten massiv erhöht. Richtwerte zur Parameterwahl für die Berechnung der SOM sind in der Literatur nur selten zu finden (Skupin und Esperbé 2011). Die Parameter richten sich in der vorliegenden Berechnung nach Lacayo-Emeri (2011). Die erste Trainingsphase der SOM wird mit einer *run length* von 60'000, einer *learning rate* von 0.04 und einer *neighborhood size* von 100 berechnet. Skupin und Esperbé (2011) erreichen mit einem ähnlich grossen Datensatz und gleichem *run length*-Parameter gute Resultate. Die *learning rate* regelt das Gleichgewicht der Spezialisierung in der SOM: eine hohe *learning rate* führt zu Bereiche in der SOM, die überspezialisiert sind. Eine tiefe *learning rate* führt hingegen zu diffusen Strukturen in der SOM. Mit einer hohen *neighborhood size* wird erreicht, dass sich in der ersten Trainingsphase globale Strukturen herausbilden, die sich in der zweiten Trainingsphase mit einer kleineren *neighborhood size* verfeinern lassen. Die Parameter für die zweite Trainingsphase lauten demnach: 100'000 für die *run length*, 0.03 für die *learning rate* und 10 für die *neighborhood size*. Die gewählten Parameter werden im Rahmen dieser Arbeit nicht evaluiert. Die resultierende SOM wird jedoch hinsichtlich ihres Zwecks (explorative Analyse der inneren Strukturen der Funktionen) auf ihre Angemessenheit überprüft. Die SOM wird nachträglich wie im Kapitel 3.2 geometrisch verzerrt, um der Distanz-Ähnlichkeitsmetapher zu entsprechen.

Die Klassierung des bimodalen Netzwerkes mittels dem Blondel-Algorithmus (Blondel et al. 2008) in NWB (NWB Team 2006) bringt einen entscheidenden Vorteil: im selben Verfahren werden Themen und Artikel thematisch kohärent gruppiert. Somit ist die Beziehung zwischen den Artikeln und den Funktionen als auch die Beziehung zwischen den Themen und den Funktionen eindeutig und kohärent.

In der SOM ergeben sich drei geometrische Auflösungen:

1. Die höchste Auflösung stellen die einzelnen Akteure dar. Diese sind jedoch wegen der grossen Anzahl (225'701) nicht darstellbar.
2. Die Neuronen sind die zweithöchste Auflösungsstufe. Für jedes Neuron wird die relative Häufigkeit der Funktionen ermittelt und der Funktion mit der maximalen relativen Häufigkeit zugewiesen (Formel 6). Die Funktionen werden mit Farbtönen und die Häufigkeiten mit Helligkeiten dargestellt. Dadurch lassen sich

die Kohäsion der Funktionen und deren Ähnlichkeit hervorheben. Funktionen, die sich auf wenige naheliegende Neuronen konzentrieren, sind kohärenter. Funktionen, die im topologischen Raum nahe zueinander sind, sind ähnlicher.

Formel 6: Maximale relative Häufigkeit der Funktionen pro Neuron.

$$\max_i \frac{af_{ni}}{af_i \times af_n}$$

wobei af_{ni} die Häufigkeit der Artikel in Neuron n , welche zur Funktion i gehören, af_i die Häufigkeit an Neuronen in Funktion i und af_n die Häufigkeit an Artikeln im Neuron n sind.

3. Die tiefste Auflösung ergibt sich durch die Wechselwirkung zwischen Neuronen in denen Akteure vorkommen und solchen, in denen keine vorkommen. Metaphorisch entsprechen die Gruppen an angrenzenden Neuronen Inselregionen umgeben von leerem Raum. Nur auf den Inselregionen kommen Akteure vor. Die verwendete Primitive ist das Aggregat. Für jede Inselregion lässt sich die Grösse, gemessen an der Anzahl Akteure und die funktionale Zusammensetzung bestimmen. Dargestellt werden sie mittels Kreisdiagramme.

Auf Attributsebene lassen sich auch drei Auflösungsebenen identifizieren. Neben den acht Funktionen, gibt es elf Subfunktionen. Zwei der acht Funktionen teilen sich auf einem tieferen Aggregationsniveau in zwei bzw. drei Subfunktionen auf. Die höchste Auflösung nehmen die 83 latenten Themen ein. Die Ausprägung jedes Themas pro Neuron wird bei der SOM-Generierung berechnet. Bei der Darstellung wird jedes Neuron dem Thema mit dem maximalen Wert zugesprochen. Dabei entstehen Gruppen angrenzender Neuronen (thematische Regionen) in denen ein latentes Thema besonders ausgeprägt ist.

Dank dem Ebenenprinzip lassen sich mehrere Informationen gleichzeitig darstellen und in Bezug zueinander setzen. Die Inselregionen können über die latenten Themen gelegt werden, wodurch der Bezug Inselregion - latentes Thema intuitiv ersichtlich wird. Inselregionen, die sich in einer thematischen Region befinden, haben einen entsprechenden thematischen Schwerpunkt.

Wegen des semantischen Charakters der Daten ist die Hervorhebung ihres Inhalts besonders relevant. Die Artikel- und die Kategorientitel bilden dabei den Textkorpus. Der Text wird vor der weiteren Analyse bereinigt. Pronomen, Artikel, Präpositionen, Konjunktionen, Zahlen, sowie die Konjugation der Verben „be“ und „have“ werden gelöscht. Zusätzlich werden seltene Begriffe (Häufigkeit geringer als fünf) nicht weiter berücksichtigt. Dies ergibt eine Liste von 18'432 Begriffen.

Für jeden Begriff wird mit der tf-idf-Methode (Manning et al. 2008) (Formel 7) die Relevanz bestimmt. Die Relevanz ist skalenabhängig; somit lässt sich für jeden Begriff auf jeder geometrischen und attributalen Auflösungsebene die Relevanz berechnen. Die Dokumente entsprechen je nach Auflösung den Neuronen, den Inselregionen, den la-

tenten Themen, den Subfunktionen oder den Funktionen. Die Dokumentlänge entspricht der Anzahl Artikel und Kategorien pro Einheit. Bei der Berechnung der idf_t werden die Daten gefiltert: nur Dokumente, in denen die absolute Häufigkeit eines Begriffes höher als eins und die relative Häufigkeit einen auflösungsabhängigen Grenzwert überschreitet (gemessen an der Gesamthäufigkeit des Begriffes im Korpus), zählen als df_t . Der Grenzwert beträgt 1.00% für die Relevanz auf Funktionsebene, 0.66% auf Subfunktionsebene und 0.00% für alle detaillierteren Auflösungen.

Formel 7: Relevanz eines Begriffes gemessen mit der tf-idf-Methode (Manning et al. 2008)

$$tf-idf_{t,d} = tf_{t,d} \times idf_t = \frac{tf_{t,d}}{dl_d} \times \log_{10} \frac{N}{df_t}$$

wo $tf_{t,d}$ die Häufigkeit des Begriffes t im Dokument d , dl_d die Dokumentlänge von d , N die Anzahl Dokumente im Korpus, df_t die Anzahl Dokumente, in denen der Begriff t vorkommt sind.

Ein Begriff hat einen hohen Wert, wenn er oft in einem Dokument und gleichzeitig in sehr wenigen anderen Dokumenten vorkommt. Bei allgemeinen Begriffen, die für keine Gruppe spezifisch sind, fällt der Wert aufgrund der inversen Dokumenthäufigkeit (idf) klein aus, obwohl sie sehr häufig sind.

Einschätzung der Themengrösse

Wie im Kapitel Datengrundlage erwähnt finden unterschiedliche Themen unterschiedliche Beachtung in Wikipedia, obwohl ein Thema nicht zwingend relevanter ist als ein anderes. Diese Verzerrung kommt insbesondere bei der Einschätzung der Grösse der Funktionen zum Tragen. So könnte der Grund für eine grosse Anzahl Artikel in der Funktion Politik nicht in der Wichtigkeit dieses Themas, sondern in einer durch die *crowdsourcing* bedingte Verzerrung liegen. Aus diesem Grund werden die vorliegenden methodischen Schritte zur Bestimmung der Funktionen mit einem Datensatz aus zufällig ausgewählten Artikeln (Zufallsdatensatz) wiederholt. Der Datensatz umfasst dieselbe Anzahl Artikel (234'059 Artikel).

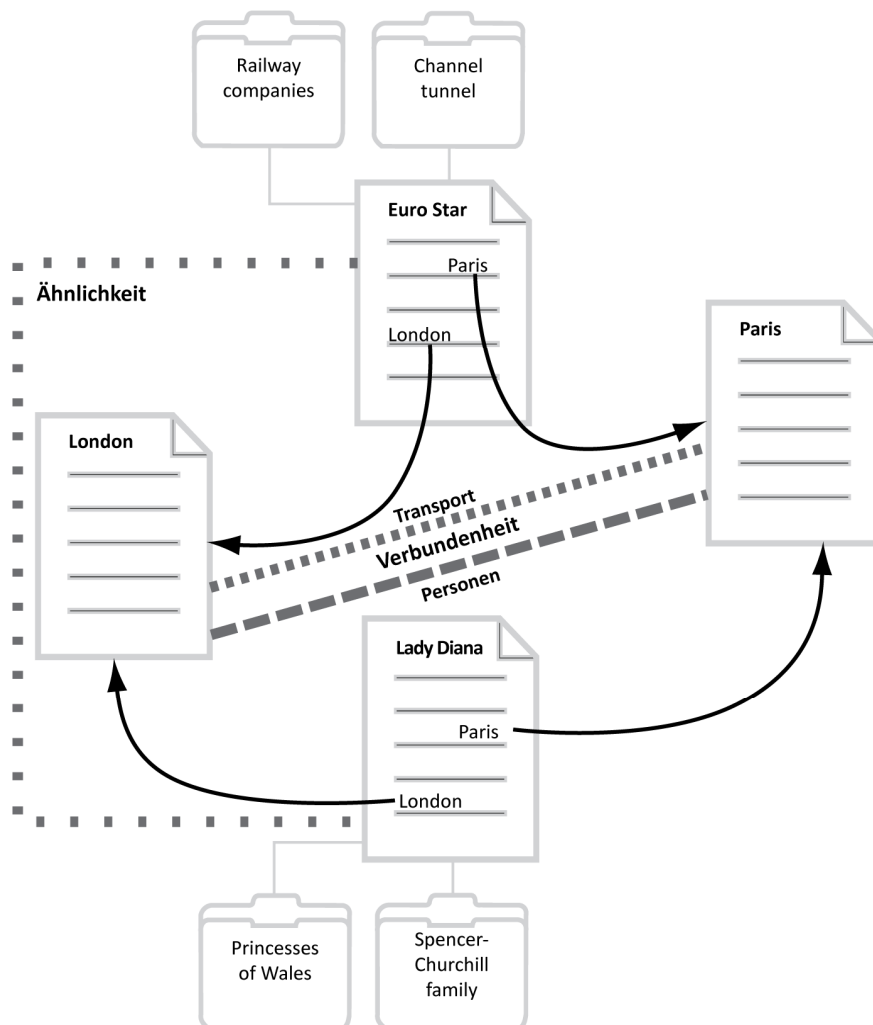
Da die Anzahl Funktionen (=Subgruppen nach Blondel) im Zufallsdatensatz und im thematischen Datensatz ungleich sind, ist der Vergleich zwischen diesen beiden schwierig. Aus diesem Grund stützt sich der Vergleich nicht auf die Funktionen, sondern auf die 83 Themen. Diese sind in beiden Datensätzen gegeben. Jeder Artikel wird demjenigen Thema zugewiesen, zu dem er die stärkste Beziehung hat. Dadurch ergibt sich ein bimodales Netzwerk zwischen den 225'701 Artikeln und den 83 Themen. Schliesslich wird für beide Datensätze die relative Artikelhäufigkeit pro Thema ermittelt.

Aus der Ermittlung der Funktionen im thematischen Datensatz ist bekannt, welche Themen zu welchen Funktionen gehören. Basierend auf dieser Gruppierung werden die Themen zusammengefasst und die relative Artikelhäufigkeit pro Funktion für den thematischen und den Zufallsdatensatz ermittelt. Da die Anzahl Funktionen somit konstant ist, ist ein direkter Vergleich der thematischen Schwerpunkte in den Datensätzen möglich.

5.4 Analyse des funktionalen Städtetetzes

In diesem Kapitel werden die Modellierung des Städtetetzes und die definierten Funktionen zusammengeführt, um ein funktionales Netzwerk zu eruieren. Die Modellierung der funktionalen Städtetetze erfolgt analog zur Modellierung des allgemeinen Städtetetzes unter Berücksichtigung von jeweils nur denjenigen Artikeln, die zu einem ausgewählten Thema gehören (Abbildung 16). Dadurch ergeben sich acht funktionale Netzwerke. Die funktionalen Netze beschränken sich auf die 95 Hubs.

Abbildung 16: Einbezug der Artikel, Hyperlinks und Kategorien für die Bestimmung von funktionalen Beziehungen zwischen zwei Städten



Bei der Analyse des funktionalen Netzwerkes stehen zwei Aspekte im Zentrum des Interesses. Einerseits werden ein funktional differenziertes Netzwerk und drei funktionale Netzwerke modelliert und untersucht. Somit wird insbesondere die funktionale Stellung der Städte und die Struktur der funktionalen Netzwerke untersucht. Andererseits wird die Ähnlichkeit zwischen den Städten aufgrund der funktionalen Verflechtungen eruiert und analysiert.

Die Analyse der funktionalen Netzwerke erfolgt analog zur Methodik für das allgemeine Städtetz (Kapitel 5.1). Zentralitätsmasse, Subgruppen und Rollen sind dabei die wichtigsten Merkmale. Die Visualisierungen basieren auf Netzwerkdarstellungen mit dem GEM-Algorithmus (Frick et al. 1994) und stützen sich dabei auf die Distanz-Verbindenheitsmetapher. Im funktional differenzierten Netzwerk wird für jede Beziehung im MST die Funktion mit der überdurchschnittlichsten Häufigkeit ermittelt und hervorgehoben. Die Verwendung der graphischen Variablen folgt der definierten Grammatik (Tabelle 13).

Der Vergleich der Städte unter Berücksichtigung aller Funktionen basiert auf den Anteilen jeder Funktion am gewichteten *degree* pro Stadt. Die bimodale Städte-Funktionen-Matrix wird mit dem SOM Analyst (Lacayo-Emeri 2011b) in den topologischen Raum projiziert. Dieser besteht aus 10x10 Neuronen, in den die 95 Städte projiziert werden. Dabei werden mehr die Dimensionsreduktions- als die Clusteringeigenschaften der Methoden genutzt.

Jede Funktion lässt sich in der SOM als Komponentenraum mit Farbton und Helligkeit darstellen. Somit ist der Verlauf jeder Funktion im topologischen Raum ersichtlich. Bei einer gleichzeitigen Darstellung aller Funktionen werden die Neuronen aufgrund der Ausprägung der Funktionen (bimodale Neuronen-Funktionen-Matrix) mittels Blondel-Algorithmus klassiert. Damit entstehen Gruppen von Neuronen mit ähnlichem Funktionenmix. Auf der Ebene der BMU werden die Städte mit ihren Eigenschaften dargestellt. Die Visualisierung gibt die Container-Metapher optimal wieder: Unterschiedliche Funktionen, die als Polygone dargestellt sind, enthalten einzelne Punkte, welche die Städte darstellen. Städte, die besonders stark von einer Funktion geprägt sind, sind auch in dieser funktionalen Region enthalten. Mit dieser Visualisierung werden der Zusammenhang zwischen Städten und Funktionen einerseits und die Ähnlichkeit der Städte aufgrund dieser Funktionen andererseits besonders ersichtlich.

5.5 Evaluation

Bei der Beantwortung der sozialwissenschaftlichen Forschungsfrage anhand der gewählten Daten und Methoden wird von unterschiedlichen Annahmen ausgegangen. Obwohl diese Annahmen begründet sind, ist eine Evaluation des Einflusses dieser Entscheidungen auf die Resultate unabdingbar. Vier ausgewählte und prägende Aspekte der vorgestellten Methodik werden systematisch evaluiert. Als erstes stehen die Daten im Fokus. Wie oben erwähnt birgt Wikipedia unter anderem zwei mögliche Verzerrungen: erstens eine kulturelle Verzerrung, für die englischsprachige Version von Wikipedia demzufolge ein Übergewicht der Städte im englischen Sprachraum. Zweitens eine geringe Robustheit der Resultate aufgrund der Volatilität der Einträge. Der zweite Punkt betrifft die Modellierung des Netzwerkes. Dabei ist dessen Grösse, gemessen an der Anzahl gewählter Städte, ein Aspekt, der die Resultate beeinflussen kann. Noch prägender ist jedoch die Berechnung der Beziehungsstärke zwischen den Städten aufgrund der Wikipedia-Einträge.

Da die Arbeit einen sehr starken visuell-analytischen Charakter hat und insbesondere im Bereich der Netzwerkvisualisierung sehr unterschiedliche Methoden erstmals integriert, ist eine Evaluation der Dimensionsreduktionsverfahren besonders wichtig. Dabei ist die zentrale Frage, ob durch die Verräumlichung die latenten Strukturen der Daten, oder methodenbedingte Artefakten hervorgehoben werden. Der letzte Aspekt, der evaluiert wird, ist die Definition der Funktionen. Dabei stellt sich die Frage, ob die resultierenden Funktionen von der gewählten Methode abhängen oder gleiche Resultate mit anderen Methoden wiederzufinden sind.

In den folgenden Absätzen wird die Methodik zur erwähnten Evaluation erklärt. Im Kapitel 6 werden hingegen die Resultate der Evaluation dargelegt.

Datengrundlage

Um den Einfluss der Enzyklopädie Sprache zu eruieren werden die drei grössten Versionen von Wikipedia, nach der englischsprachigen, als Vergleich herbeigezogen. Dabei handelt es sich um die deutschsprachige (Stand: 2. März 2011), die französischsprachige (Stand: 10. Februar 2011) und die italienischsprachige (Stand: 3. Februar 2011) Version. Aus allen vier Datensätzen werden, nach dem im Kapitel 5.1 dargelegten Verfahren, die wichtigsten Städte, die in allen vier Versionen vorkommen, extrahiert und das Städtetz modelliert. Die vier Städtetze bestehen jeweils aus 516 Städten. Die Verbindung der 516 Städte-Artikel in der englischsprachigen Version mit den entsprechenden Artikeln in den anderen Sprachversionen erfolgt über die MySQL-Tabelle *ills* (Tabelle 14).

Vor der eigentlichen Untersuchung der Städtetze, werden folgende Kenngrößen der Enzyklopädien verglichen:

- Anzahl Artikel und Anzahl Links: Es wird erwartet, dass eine höhere Anzahl Artikel und insbesondere eine höhere Anzahl Links einen positiven Einfluss auf die Analysen hat. Die Informationen sind vollständiger und die Datenbasis robuster. Diese zwei Kennwerte werden von Zachte (2011) erfasst.
- Anzahl AutorInnen: Nach Flanagin und Metzger (2008) führt eine höhere Autorenzahl zu einer höheren Genauigkeit. Daraus lässt sich in der Tendenz ableiten, dass die Enzyklopädie mit der grössten Autorenschaft auch die genaueste ist. Die Autorenzahl wird von Zachte (2011) erfasst.
- Geographische Heterogenität der Autorenschaft: AutorInnen können in Wikipedia ein Benutzerkonto einrichten. Diese werden analog zu den Artikeln mit Kategorien organisiert. Aufgrund der Kategorie “*wikipedians by location*” (Wikipedia contributors 2011f), welche in allen vier Versionen vorkommt und die Autorenschaft nach Wohnort ordnet, wird die geographische Verteilung der Autorenschaft qualitativ eingeschätzt.
- Anzahl einkommender Links in den 516 Städten: Die Anzahl einkommender Links gibt einen Hinweis über das Prestige (Wasserman und Faust 2008) der ein-

zelen Artikel. Wenn Städte im englischsprachigen Raum in der englischen Wikipedia ein überdurchschnittlich hohes Prestige haben, wäre dies ein Indiz für eine sprach-räumliche Verzerrung.

Die Evaluation des Städtnetzes basiert einerseits auf dem englischen und andererseits auf einem multilingualen Städtnetz, welches aus dem Zusammenführen der vier einzelnen Netze entsteht. Die Kantengewichte der vier Städtnetze werden auf Werte zwischen null und 1000 normalisiert. Die neuen Kantengewichte zwischen zwei bestimmten Städten entsprechen der Summe der normalisierten Kantengewichte zwischen denselben in den vier Netzwerken. Die Aggregation der Netzwerke ermöglicht, die sprachliche Verzerrung zumindest teilweise auszugleichen.

Für das englischsprachige und das Mehrsprachennetzwerk werden folgende netzwerk-analytische Masse berechnet. Sie decken sowohl die individuelle, als auch die relationale und die Gruppen Analyseebene und somit die Hauptdimensionen des Netzwerkes ab.

- Gewichteter *degree* (Formel 3)
- Z-Wert nach Guimerà und Amaral (2005a) (Formel 5)
- Subgruppen nach Blondel (2008)
- MST (Dearholt und Schvanenveldt 1990)
- *Degree* im MST

Der MST dient als Input für die Projektion der Netzwerke im Raum mit dem GEM-Algorithmus (Frick et al. 1994). Z-Werte und *degree* lassen sich mittels Größenänderung, die Subgruppen mit Farbtonvariation der Knoten und die Kantengewichte mit Helligkeitswerten und Grösse der Kanten kennzeichnen. Der visuelle Vergleich ermöglicht eine erste Einschätzung der Übereinstimmung zwischen den Netzwerken. Ein systematischer Vergleich der Indizes erfolgt auf mathematischer Basis mit einer Korrelationsanalyse. Der Korrelationskoeffizient nach Pearson wird zwischen den Massen in den zwei Netzwerken auf individueller Ebene (Gewichteter *degree*, *degree* im MST, Z-Wert) und auf relationaler Ebene (Kantengewichte im Gesamtnetzwerk und im MST) berechnet. Der Vergleich der Subgruppen stützt sich auf die hypergeometrische Methode nach Kos und Psenicka (2000).

Volatilität der Einträge und Modellierung der Beziehungsstärke

Beide Evaluationen setzen beim bimodalen Städte-Akteur Netzwerk an. Die Volatilität der Einträge ergibt sich durch die Möglichkeit, den Inhalt der Artikel laufend zu verändern. Für die vorliegende Arbeit ist das Löschen von Verlinkungen besonders entscheidend. Aus diesem Grund werden aus dem bimodalen Städte-Akteur Netzwerk 25% der Akteure gelöscht. Die Auswahl erfolgt nach dem Zufallsprinzip.

Die Stärke der Kanten hängt massgeblich von der gewählten Normalisierung ab. In der vorliegenden Arbeit wird die Normalisierung nach Hecht und Raubal (2008) (Formel 2)

gewählt. Dafür spricht, dass diese explizit für Wikipedia entwickelt ist (Hecht und Raubal 2008). Die am naheliegendste Alternative ist eine einfache Normalisierung über die Grösse der verbindenden Artikel gemessen an der Anzahl ausgehender Links (Formel 8).

Formel 8: Einfache Berechnung des normalisierten Kantengewichts.

$$KG_N = \frac{KG}{OD}$$

wo KG_N das normalisierte, KG das nicht-normalisierte Kantengewicht und OD der *outdegree* (Anzahl ausgehender Links) ist.

Die zwei veränderten bimodalen Netzwerke werden, wie im Kapitel 5.1 erläutert, zu einem unimodalen Stadt-Stadt-Netzwerk umgerechnet und anschliessend mit dem vollständigen, unveränderten Netzwerk verglichen. Der Vergleich wird nach dem gleichen Verfahren wie oben für die Evaluation der sprachlichen Verzerrung erläutert durchgeführt.

Verräumlichung mit DrL/VxOrd

Klavans und Boyack (2006a, 2006b, 2011; Boyack et al. 2005) definieren die Arbeitsschritte für das Generieren von Verräumlichungen mit dem DrL/VxOrd-Algorithmus und evaluieren dessen Genauigkeit. Dabei postulieren sie, dass die genauesten Verräumlichungen erzeugt werden, wenn die Inputmatrix nur die 15 wichtigsten Verbindungen pro Knoten enthält. Dieser entscheidende Parameter wird jedoch nicht explizit evaluiert (Klavans und Boyack 2006a). Die folgende Evaluation greift einen methodischen Ansatz von Skupin (2003) auf: die Verräumlichungsmethode wird auf bekannte geographische Daten angewendet, um damit die vom Algorithmus erzeugten Mustern und Verzerrungen in der Visualisierung besser evaluieren und diskutieren zu können. Die vorliegende Evaluation dieses Parameters basiert somit auf den 3'532 Städten des Datensatzes und deren geographischer Lage und Distanz. Die geographische Lage jeder Stadt ist bekannt und auf einer Weltkarte darstellbar. Ein optimales Layout würde aufgrund der Distanzmatrix zwischen den Städten ein ähnliches Bild generieren, in dem die wesentlichen Strukturen erhalten bleiben.

Die Städte werden in ArcGIS aufgrund ihrer geographischen Koordinaten im Raum projiziert. Danach wird die euklidische Distanz zwischen den Städten berechnet, welche die Kantengewichte des Netzwerkes darstellen. Dabei werden nun nach dem Ansatz von Klavans und Boyack (2006a) die 15 stärksten Kanten pro Stadt ausgewählt, d.h. für jede Stadt werden die 15 nächsten Nachbarstädte berücksichtigt. Aufgrund dieses Inputs wird mit dem DrL/VxOrd-Algorithmus (Shawn et al. 2011) in NWB (NWB Team 2006) das topologische Layout berechnet. Zu Vergleichszwecken wird das Verfahren mit Netzwerken, in denen die 100 und die 500 stärksten Beziehungen pro Stadt erhalten bleiben, wiederholt.

Die resultierenden topologischen Koordinaten werden wiederum in ArcGIS eingelesen. Nun stehen drei topologische Räume dem geographischen Referenzraum gegenüber. Für jeden der vier Räume wird die Mesostruktur ermittelt: die naheliegenden Städte (Punkte) werden mit dem *Point Aggregation Tool* in ArcGis zu Regionen (Flächen) zusammengefasst. Der Grenzwert beträgt 90km (entsprechen 10mm in ArcGIS).

Die Genauigkeit wird aufgrund der folgenden skalenabhängigen Genauigkeitsmasse eruiert:

- 1 Meso-Genauigkeit: Wird aufgrund der Übereinstimmung der Regionen im topologischen und im geographischen Raum, gemessen mit dem hypergeometrischen Test für Clusterkonsistenz (Kos und Psenicka 2000), bestimmt. Der Test berechnet die Wahrscheinlichkeit dafür, dass die Gruppierung der Städte im topologischen Raum zufällig ist und nicht mit der Gruppierung im geographischen Raum übereinstimmt. Die Nullhypothese des Tests lautet somit: Die Regionen im topologischen Raum sind eine ungenügende Reproduktion der Regionen im geographischen Raum. Je tiefer der hypergeometrische Koeffizient ist, desto unwahrscheinlicher ist die Nullhypothese und genauer die Verräumlichung.
- 2 Makro-Genauigkeit: Bezieht sich auf die Distanzverhältnisse zwischen den Regionen (interregionale Distanz). Die Evaluation der Makro-Ebene stützt sich auf die Distanzen zwischen den Schwerpunkten der Regionen. Die Schwerpunkte und die gegenseitigen Distanzen werden im geographischen und im topologischen Raum berechnet und mit einer Korrelationsanalyse nach Pearson verglichen. Eine hohe Korrelation deutet auf eine hohe Genauigkeit hin.
- 3 Mirko-Genauigkeit: Die Mikro-Genauigkeit bezieht sich auf die Distanzverhältnisse innerhalb der Regionen (intraregionale Distanz). Aus drei Regionen von unterschiedlichen Dichten des geographischen Raums, in der Regel aus Amerika, Europa und Afrika, werden jeweils acht Städte ausgewählt. Bei der Auswahl wird beachtet, dass die Distanzen zwischen diesen Städten sehr unterschiedlich sind. Die 28 Distanzbeziehungen pro Gruppe (84 insgesamt pro Projektion) werden im topologischen und im geographischen Raum berechnet und mit einer Korrelationsanalyse nach Pearson verglichen. Je höher der Korrelationskoeffizient ist, desto genauer ist die Visualisierung.

Definition der Funktionen

Das Bestimmen von thematischen Schwerpunkten in einem Textkorpus stellt eine besondere Herausforderung dar, insbesondere wegen der hohen semantischen Mehrdeutigkeit. Sehr unterschiedliche Methoden sind für die Eruiierung der semantischen Ähnlichkeit und der Klassierung von Texten entwickelt worden. Die Wahl der Methode kann grundsätzlich das Schlussergebnis beeinflussen.

Die Definition der Funktionen beruht auf drei Arbeitsschritten:

- Eruierung der im Korpus latent vorhandenen Themen und Beziehung der einzelnen Artikel zu diesen Themen.
- Ableitung der Ähnlichkeiten zwischen den Artikeln aufgrund der Beziehungen zu den Themen
- Gruppierung der Artikel zu Funktionen aufgrund ihrer Ähnlichkeiten

Im Kapitel 5.3 werden die latenten Themen und die Beziehung der Artikel zu denen aus dem Kategoriensystem in Wikipedia bestimmt. Die Ähnlichkeit wird nur implizit über das bimodale Artikel-Thema-Netzwerk betrachtet, welches als Input für die Berechnung der Funktionen mit dem Blondel-Algorithmus dient.

Validiert wird insbesondere der erste Schritt mit dem *probabilistic topic model* (Steyvers und Griffiths 2007). Das *topic model* ist eine Weiterentwicklung der *latent semantic analysis* (LSA). Dabei folgt das *topic model* folgendem Grundschemata: aus dem Textkorpus werden eine bestimmte Anzahl Themen hervorgehoben. Jedes Thema wird durch gewisse Begriffe (Wörter oder Wortsequenzen) bestimmt. Aufgrund dieser Begriffe werden für jedes zu analysierende Textpaket die Wahrscheinlichkeiten für das Vorkommen der Themen berechnet. Das Resultat ist eine bimodale Matrix, in der die Textpakete zu den Themen in Bezug stehen (Steyvers und Griffiths 2007). In der vorliegenden Arbeit sind die Artikel und die verbundenen Kategorientitel die zu analysierenden Textpakete. Die Analyse erfolgt mit der *text visualization toolbox* (Rebich Hespanha und Hespanha 2011). Um die Vergleichbarkeit mit der Klassierung anhand der Wikipedia-Themen zu erhöhen, werden die Parameter gleich gewählt. Im Spezifischen wird die Anzahl Themen auf 83 gesetzt.

Die Funktionen werden analog zum Kapitel 5.3 aus der bimodalen Artikel-Themen-Matrix mit dem Blondel-Algorithmus bestimmt. Mit dem *topic modelling* werden die Akteure zu 59 Funktionen gruppiert. Der grosse Unterschied in der Funktionszahl erschwert einen Vergleich der zwei Lösungen. Aus diesem Grund werden die 59 Funktionen aus dem *topic modelling* zu acht „Überfunktionen“ zusammengefasst, die den acht Funktionen aus Wikipedia entsprechen. Jede der 59 Funktionen wird der Funktion nach Wikipedia zugewiesen zu der die grösste Überlappung besteht. Die grösste Überlappung entspricht der Intersektion zwischen den beiden Funktionslösungen mit der grössten relativen Übereinstimmung (Formel 9).

Formel 9: Relative Übereinstimmung einer Funktion nach Wikipedia (FWI) und nach *topic modelling* (FTM)

$$\text{relative Übereinstimmung} = \frac{af_{F_i, WI \cap F_i, TM}}{af_{F_i, WI} \times af_{F_i, TM}}$$

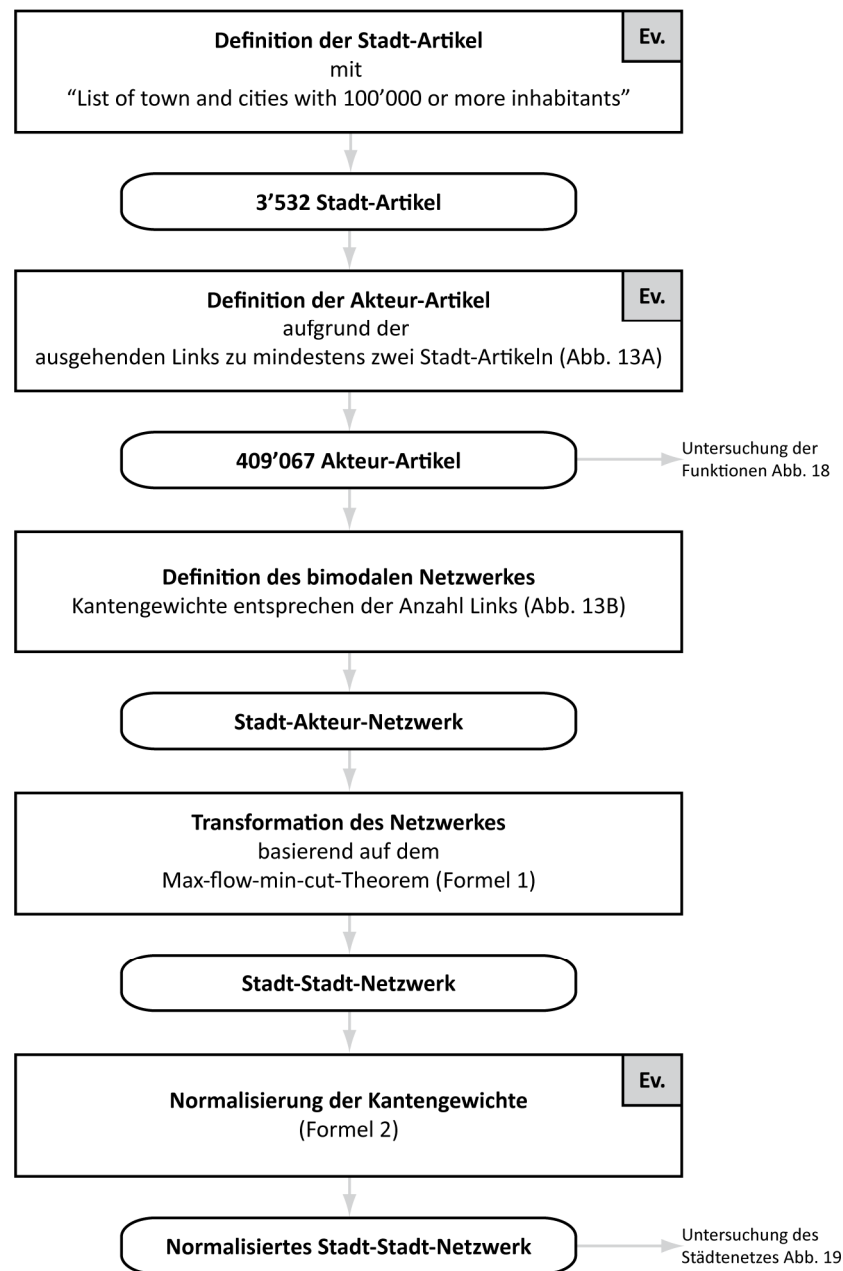
wobei $af_{F_i, WI}$ der Anzahl Akteure in der Funktion i nach Wikipedia, $af_{F_i, TM}$ der Anzahl Akteure in der Funktion i nach *topic modelling* und $af_{F_i, WI \cap F_i, TM}$ der Anzahl Akteure in der Intersektion der beiden Funktionen entsprechen.

Die acht Funktionen werden mit den acht „Überfunktionen“ inhaltlich, visuell und mathematisch verglichen. Mathematisch werden die zwei Lösungen mit dem Kappa Koeffizient (Cohen 1960) und dem hypergeometrischen Test für Clusterkonsistenz (Kos und Psenicka 2000) verglichen. Je höher die Übereinstimmung der Cluster ist, desto höher die Robustheit der Klassierung. Für jede Funktion werden zudem der *error of omission* und der *error of commission* berechnet. Die Funktionsdefinition mit Wikipedia dient als Referenz. Der *error of omission* umfasst die Elemente einer Funktion nach Wikipedia, welche nicht als solche in der Funktionsdefinition nach dem *topic modelling* erkannt werden. Der *error of commission* umfasst die Elemente, welche im *topic modelling* zusätzlich einer Funktion zugerechnet werden.

5.6 Fazit

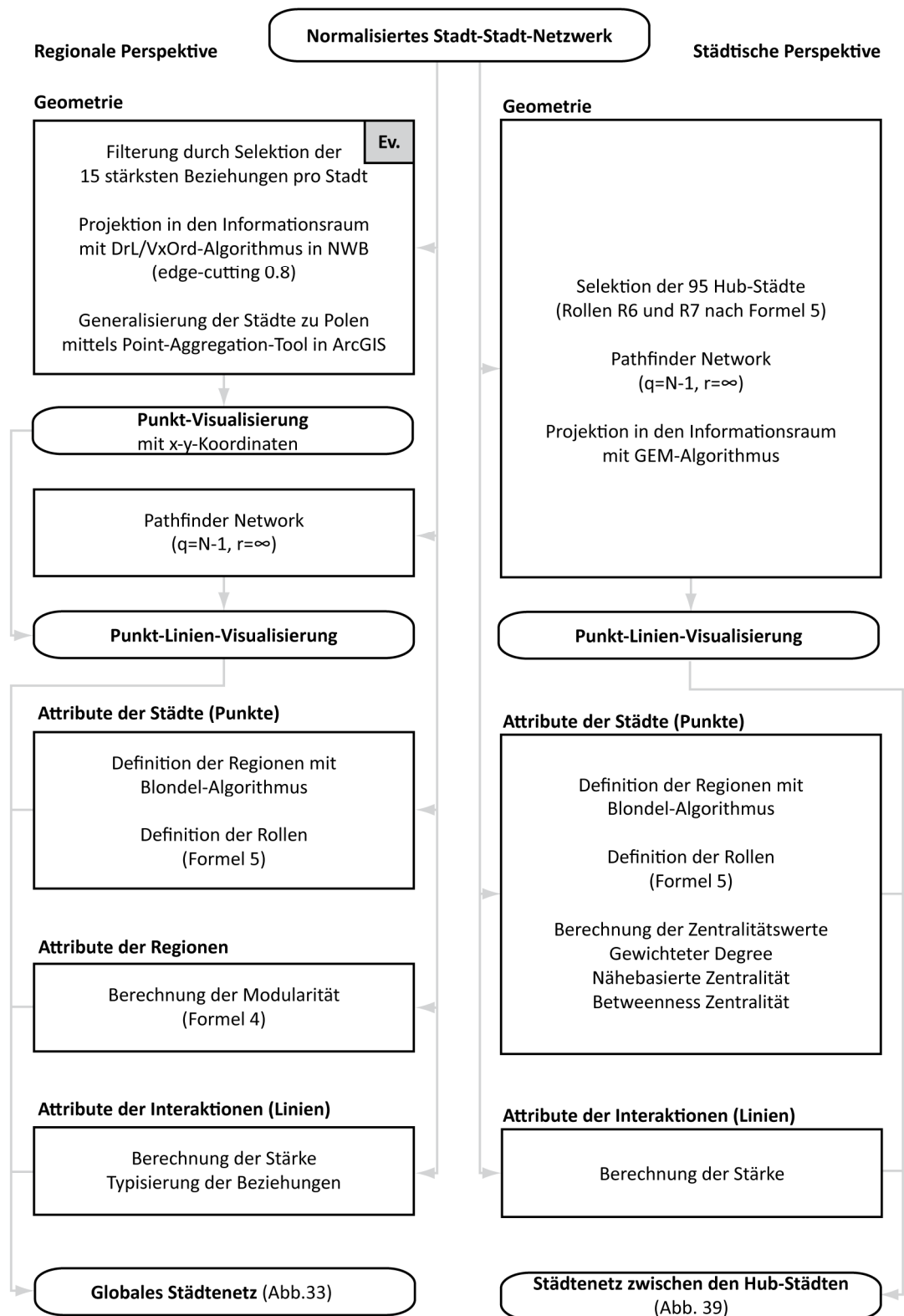
Die Modellierung des Städtetetzes und der Funktionen stellen die zwei Grundpfeiler der Methodik dieser Arbeit dar. Die folgenden drei Flussdiagramme fassen den in Kapitel 5 beschriebenen methodischen Arbeitsablauf zusammen. Abbildung 17 hebt insbesondere die Transformation der Wikipedia-Daten zu einem Städtetetz hervor. Dabei bildet die Selektion der Städte den ersten Schritt des Arbeitsablaufs. Das globale Städtetetz in der Form eines normalisierten Stadt-Stadt-Netzwerks stellt das Endresultat dar. Die methodischen Einzelheiten und die Parameter werden in Kapitel 5.1 ausführlich beschrieben. Die Abkürzung „Ev.“ in den Abbildungen zeigen auf, dass die entsprechenden methodischen Schritte evaluiert werden.

Abbildung 17: Arbeitsschritte für die Transformation der Rohdaten (Wikipedia) in ein Städtenetz



Das normalisierte Stadt-Stadt-Netzwerk stellt wiederum der Ausgangspunkt für die Visualisierung und die Analyse des globalen Städtenetzes dar. Der generische methodische Ablauf aus Kapitel 3.2 und die in Kapitel 5.2 detailliert beschriebene methodische Vorgehensweise zur visuell-analytischen Untersuchung sind in Abbildung 18 zusammengefasst. Dabei stehen zwei Arbeitsabläufe zur Verfügung. Zwei verschiedene Perspektiven, die zu unterschiedlichen Visualisierungen führen, werden dabei unterschieden. Einerseits eine Visualisierung, die die regionale Perspektive hervorhebt. Diese Resultate werden in Kapitel 7.1 – Regionale Perspektive vorgestellt. Andererseits wird in Kapitel 7.1 – Städtische Perspektive eine Visualisierung präsentiert, welche die städtische Perspektive in den Vordergrund stellt.

Abbildung 18: Arbeitsschritte zur Visualisierung und Analyse des globalen Städtenetzes

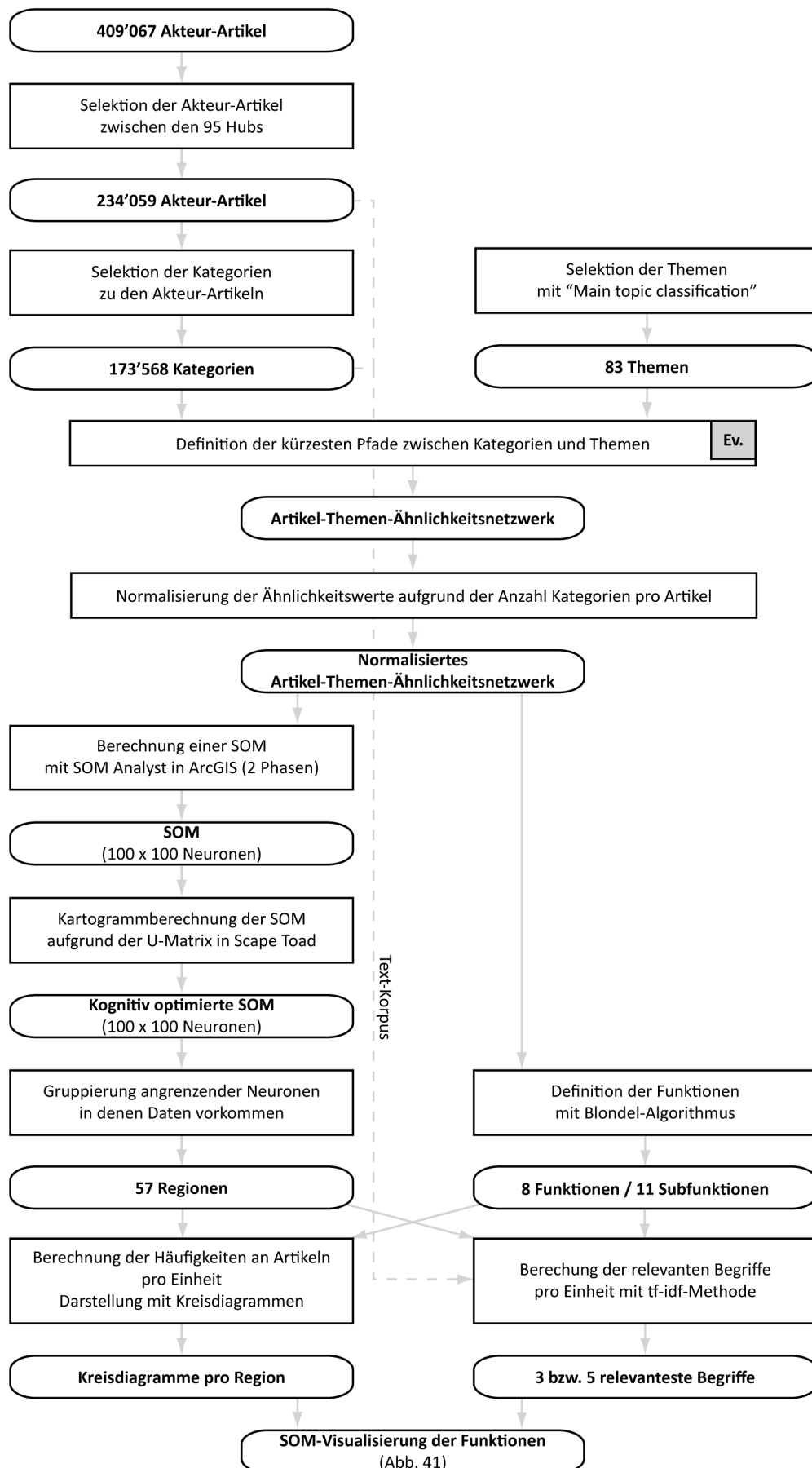


Die Visualisierung und Analyse der Funktionen beginnt bei der Liste der Akteur-Artikel. Neben dieser Liste der Akteur-Artikel stützen sich die Definition und die folgende visuell-analytische Untersuchung der Funktionen auf das Kategoriensystem. Der detaillierte Arbeitsablauf ist in Abbildung 19 illustriert und in Kapitel 5.3 ausführlich beschrieben.

Der Prozess mündet in SOM-Visualisierungen (z.B. Abbildung 41), welche die visuelle Grundlage für die Resultate in Kapitel 7.2 bilden.

Die Modellierung der funktionalen Städtenetze (Kapitel 5.4) richtet sich nach demselben Arbeitsablauf wie für das allgemeine Städtenetz (Abbildung 17, Abbildung 18).

Abbildung 19: Arbeitsschritte zur Visualisierung und Analyse der Funktionen



6 Methodenevaluation

Dieses Kapitel legt die Resultate der im Kapitel 5.5 besprochenen methodischen Ansätze zur Evaluation der Methodik und der Datengrundlage dar. So widmet sich der erste Teil der Evaluation der Datengrundlage, insbesondere der Einschätzung der Robustheit der Resultate gegenüber einer Veränderung der Sprache der verwendeten Enzyklopädie. Des Weiteren wird die Genauigkeit der Strukturen und Muster, welche mit dem DrL/VxOrd-Algorithmus graphisch erzeugt werden, überprüft. Im letzten Teil werden die Definitionen der Funktionen anhand zwei unterschiedlicher Ansätze verglichen, um den Einfluss der Methode auf die Resultate zu prüfen.

6.1 Datengrundlage

Aus Kapitel 4 ist bekannt, dass die Genauigkeit und die Volatilität der Links zwischen den Artikeln (Denning et al. 2005) sowie die räumliche Verzerrung des Inhaltes (Flanagin und Metzger 2008) drei mögliche Risiken von Wikipedia als Datengrundlage für die vorliegende Arbeit darstellen können. Die räumliche Verzerrung spiegelt sich in einer höheren Anzahl und Ausführlichkeit der Artikel über Orte, Themen, Personen und Ereignisse wider, die zum Umfeld der AutorInnen gehören. Für eine Analyse mit räumlichem Fokus ist insbesondere dieser letzte Punkt entscheidend. Die genannten Risiken sind in den 282 Sprachversionen von Wikipedia vorhanden, jedoch in unterschiedlichem Ausmass. Eine grössere Autorenschaft, eine stärkere Heterogenität ihrer Wohnstandorte (Flanagin und Metzger 2008) sowie eine grössere Datenbasis (Anzahl Artikel und Links) führen tendenziell zu einer genaueren und räumlich ausgeglicheneren

Datenbasis. Tabelle 16 zeigt die Anzahl AutorInnen, die geographische Heterogenität ihrer Wohnstandorte, sowie die Anzahl Artikel und Links für die vier umfangreichsten Sprachversionen von Wikipedia. Basierend auf diesen Kennzahlen gilt die englischsprachige Wikipedia als genauer und räumlich ausgeglichener als die anderen Sprachversionen. Die Anzahl AutorInnen ist etwa sechs Mal und die Anzahl Artikel und Links etwa drei Mal höher als in der zweitgrössten deutschsprachigen Version.

Tabelle 16: Einfache Kennzahlen der vier grössten Versionen von Wikipedia (englischsprachig, deutschsprachig, französischsprachig und italienischsprachig) nach Zachte (2011)

Kennzahlen	Englisch	Deutsch	Französisch	Italienisch
AutorInnen (>10 Einträge)	652'000	103'000	63'000	36'000
Geographische Heterogenität der Autorenschaft	<i>hoch</i>	<i>tief</i>	<i>tief</i>	<i>tief</i>
Artikel	3.5 Mio	1.2 Mio	1.1 Mio	0.8 Mio
Links	78 Mio	25 Mio	22 Mio	18 Mio

Die englischsprachige Version wird nun insbesondere hinsichtlich ihrer potentiellen räumlichen Verzerrung evaluiert. Die Evaluation basiert auf dem Vergleich zwischen einem Datensatz aus der englischsprachigen Wikipedia und einem äquivalenten mehrsprachigen Datensatz, der sich aus der englisch-, deutsch-, französisch- und der italienischsprachigen Version zusammensetzt. Der mehrsprachige Datensatz dient somit als Referenz.

Die zwei Datensätze entsprechen zwei Städtenetzen mit den gleichen 516 Städten, modelliert nach der in Kapitel 5.1 diskutierten Methodik. Die MST der zwei Netzwerke sind in Abbildung 20 (englischsprachiger Datensatz) und Abbildung 21 (Referenzdatensatz) dargestellt. Mit Farbtönen ist die regionale Organisation⁶ und mit der Grösse der Punktsymbole die lokale Zentralität der Städte hervorgehoben. Die Helligkeit und die Grösse der Linien stellt die Beziehungsstärke zwischen den Städten dar.

⁶ Die relationalen Regionen werden im Text mit KAPITÄLCHEN geschrieben.

Abbildung 20: Städtenetz basierend auf der englischsprachigen Wikipedia. Dargestellt sind 127 von 516 Städten und der MST; periphere Städte sind nicht dargestellt

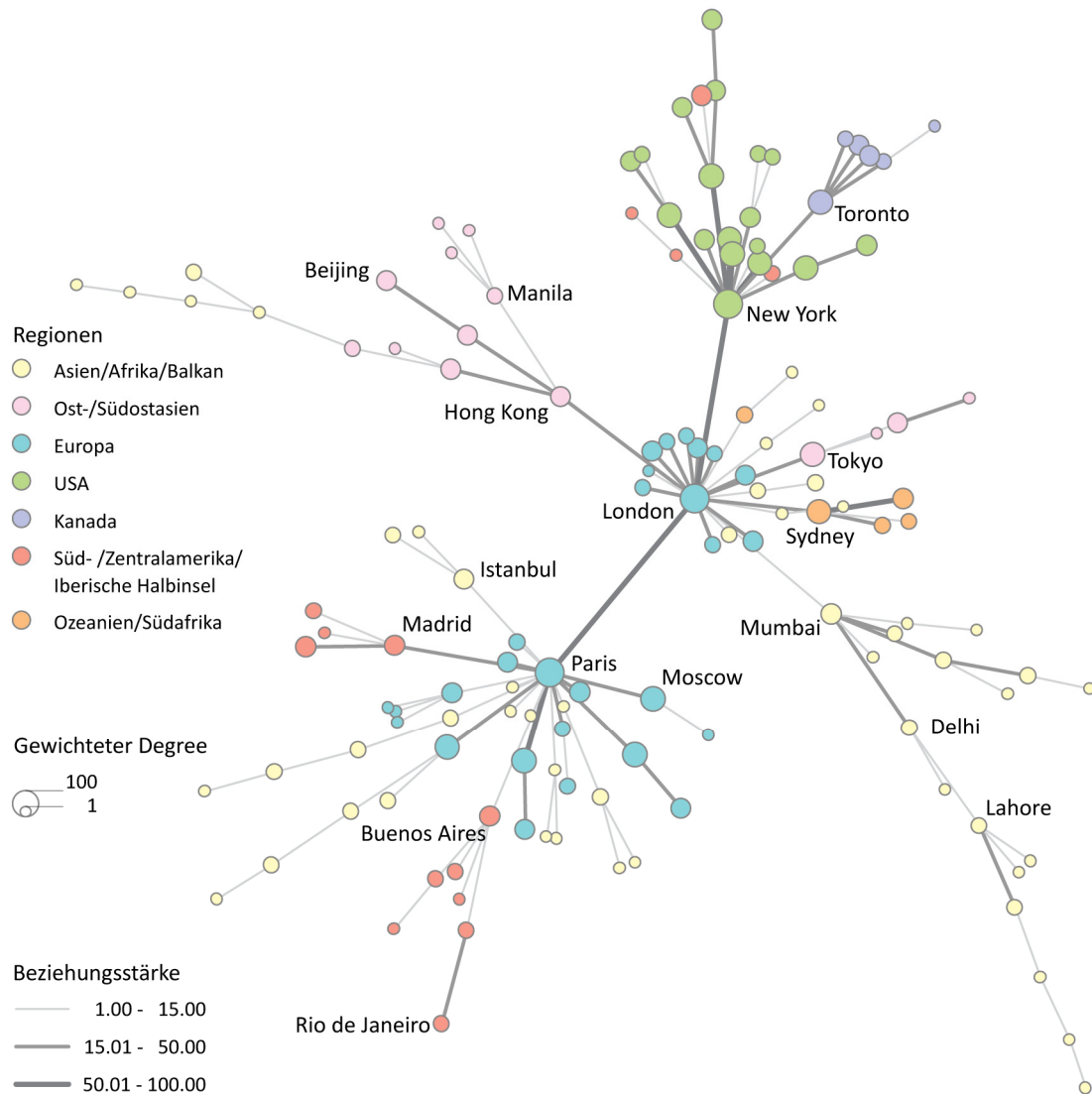
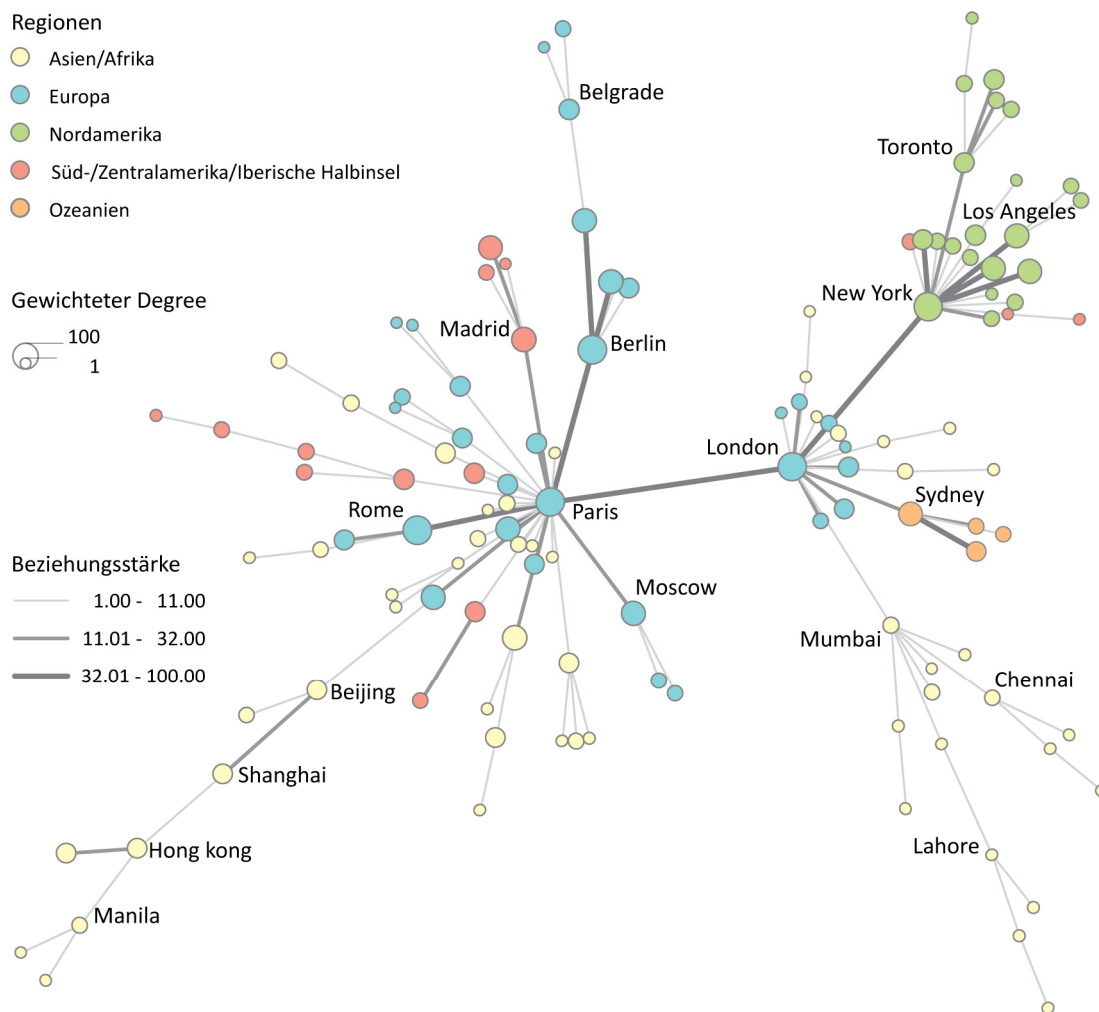


Abbildung 21: Stdtenetz basierend auf der englischen, deutschen, franzsischen und italienischen Wikipedia. Dargestellt sind 138 von 516 Stdte und der MST; periphere Stdte sind nicht dargestellt



Die hnlichkeiten bzw. die Differenzen zwischen dem englischsprachigen Datensatz (Abbildung 20) und dem Referenzdatensatz (Abbildung 21) werden mit dem Vergleich zwischen folgenden Netzwerkeigenschaften ermittelt:

- 1 drei Zentralittsmassen (Gewichteter *degree*, *degree* im MST und Z-Wert),
- 2 den Kanten und Kantengewichten im Gesamtnetzwerk und im MST und
- 3 der regionalen Organisation.

Der Vergleich erfolgt einerseits visuell, durch die Gegenberstellung der hervorgehobenen Strukturen in Abbildung 20 und Abbildung 21 und andererseits anhand Korrelationsanalysen zwischen den erwhnten Messwerten im Referenz- und im englischsprachigen Netzwerk (Tabelle 17). Die regionalen Organisationen werden mittels einer Vergleichsmatrix (Tabelle 18) gegenbergestellt.

Tabelle 17: Korrelationsanalyse zwischen dem Referenz- und dem englischsprachigen Städtetnetz aufgrund ausgewählter Netzwerkmasse

Masse	r (Korrelationskoeffizient nach Pearson)
Gewichteter <i>degree</i>	0.923
<i>Degree</i> im MST	0.961
Z-Wert	0.885
Kantengewichte im Gesamtnetzwerk	0.880
Kantengewichte im MST	0.961

Die mathematisch bestimmte Ähnlichkeit zwischen den Kantengewichten in den beiden Gesamtnetzwerken ist, mit einem Korrelationswert von $r=0.880$, sehr gross. Noch höher ist die Ähnlichkeit der zwei MST ($r=0.961$). Diese Übereinstimmung ist auch visuell erkennbar, die zwei Netzwerke sind sich sehr ähnlich: London, Paris und New York liegen im Zentrum mit starken Beziehungen zueinander. Die NORDAMERIKANISCHEN Städte beziehen sich mit teilweise starken Beziehungen auf New York, die SÜD-/ZENTRALAMERIKANISCHEN/IBERISCHEN Städte mit mehrheitlich schwachen Beziehungen auf Paris und New York, die EUROPÄISCHEN Städte mit meistens starken Beziehungen auf Paris sowie London und die OZEANISCHEN/SÜDAFRIKANISCHEN Städte mit starken Beziehungen ausschliesslich auf London. Die deutlichsten Unterschiede bestehen in den Beziehungen zwischen den asiatischen Städten in der Region ASIEN/AFRIKA/BALKAN bzw. ASIEN/AFRIKA: diese sind im englischsprachigen Netzwerk (Abbildung 20) tendenziell stärker. Besonders ausgeprägt sind die Unterschiede auf dem Netzwerkast, der Peking und Hong Kong mit Manila verbindet und die OST-/SÜDOSTASIATISCHEN Städte umfasst. Im englischsprachigen Netzwerk bezieht sich dieser Ast auf London und teilt sich auf drei Nebenäste auf. Im Referenzdatensatz (Abbildung 21) beziehen sich diese Städte auf Paris und sind kettenartig verbunden.

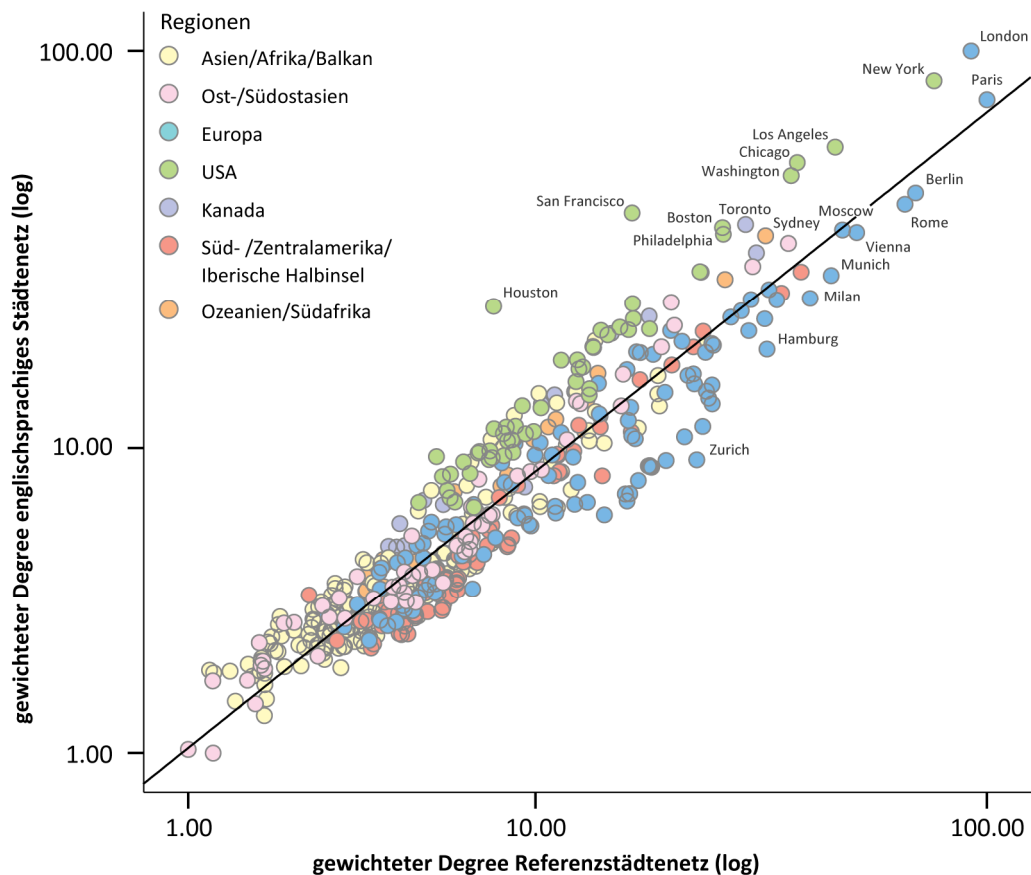
Ähnliche Netzwerkstrukturen führen zwangsläufig zu ähnlichen Zentralitätswerten. Die Korrelationswerte (Tabelle 17) zwischen den gewichteten *degree* ($r=0.923$), den *degree* im MST ($r=0.961$) und den Z-Werten⁷ ($r=0.885$) sind erwartungsgemäss sehr hoch. Die visuelle Analyse untermauert die Korrelationswerte der ersten zwei Zentralitätsmasse. London, Paris und New York sind in beiden Netzwerken die zentralsten Städte. Danach folgen vor allem weitere EUROPÄISCHE und NORDAMERIKANISCHE Städte.

Ein genauerer Vergleich der drei Zentralitätsmasse in den beiden Netzwerken bringt trotz grosser Übereinstimmung gewisse räumliche Verzerrungen zum Vorschein. Der Vergleich wird in Abbildungen 22, 23 und 24 anhand von drei Scatterplots hervorgehoben. Die x-Achse stellt dabei jeweils die normalisierten und logarithmierten Werte im mehrsprachigen Referenznetzwerk und die y-Achse die normalisierten und logarithmierten Werte im englischsprachigen Städtetnetz dar. Die Punktsymbole entsprechen den Städten, die Farbtöne den Regionen aus Abbildung 20 und Abbildung 21 und die Linien

⁷ Der Z-wert ist in Abbildung 20 und Abbildung 21 nicht dargestellt.

den linearen Regressionsgeraden. Städte, die sich oberhalb der Regressionslinie befinden, haben einen höheren Zentralitätswert im englischsprachigen Netz, als er aufgrund des Referenzwerts zu erwarten wäre. Bei Städten unterhalb der Linie ist dieser Wert tiefer als erwartet. Je grösser die vertikale Abweichung der Punkte zur Regressionsgerade ist, desto grösser ist die numerische Abweichung vom erwarteten Wert.

Abbildung 22: Scatterplot mit Regressionsgerade zwischen dem gewichteten *degree* im englischsprachigen und im Referenzstädtenetz



Der gewichtete *degree* (Abbildung 22) ist im englischsprachigen Netzwerk für viele US-AMERIKANISCHE (grün) und KANADISCHE (violett) Städte (z.B. New York, Los Angeles, Boston, Toronto, San Francisco, Houston) sowie London höher als erwartet und für mehrere EUROPÄISCHE (blau) Städte (z.B. Berlin, Rom, Mailand, Hamburg, Zürich) tiefer als erwartet. Dies ist ein klares Zeichen für eine räumliche Verzerrung in der Datenbasis: die Zentralität, gemessen mit dem gewichteten *degree* von Städten des angelsächsischen Sprachraums, wird tendenziell überschätzt. Kein systematisches Muster ist hingegen in den Abweichungen des *degrees* im MST (Abbildung 23) und des Z-Wertes (Abbildung 24) erkennbar. Die räumliche Verzerrung in der Datenbasis erscheint nicht mehr. In Abbildung 23 wird z.B. der *degree* im MST von London, Toronto und Buenos Aires über- und von Paris, New York und Mexiko City unterschätzt. Die Abweichungen sind generell geringer als für den gewichteten *degree* (Abbildung 22). Die Abweichungen der Z-Werte (Abbildung 24) sind wiederum grösser. London, Mumbai, Bukarest und Zagreb werden im englischsprachigen Städtenetz überschätzt und z.B. Peking, Bangkok,

Jakarta, Johannesburg und Colombo unterschätzt. Die Z-Werte sind nicht nur von der Zentralität einer einzigen Stadt, sondern auch von der regionalen Organisation beeinflusst. Da diese zwischen dem englischsprachigen und dem Referenznetzwerk variieren, werden dadurch auch die Z-Werte beeinflusst. Dies erklärt die teilweise grossen Abweichungen für die OST-/SÜDOSTASIATISCHEN, ASIATISCH/AFRIKANISCH/BALKANISCHEN und OZEANISCH/SÜDAFRIKANISCHEN Städte. Die systematische räumliche Verzerrung ist im Vergleich zu Abbildung 22 geringer.

Abbildung 23: Scatterplot mit Regressionsgerade zwischen dem Z-Wert im englischsprachigen und Referenzstädtenetz

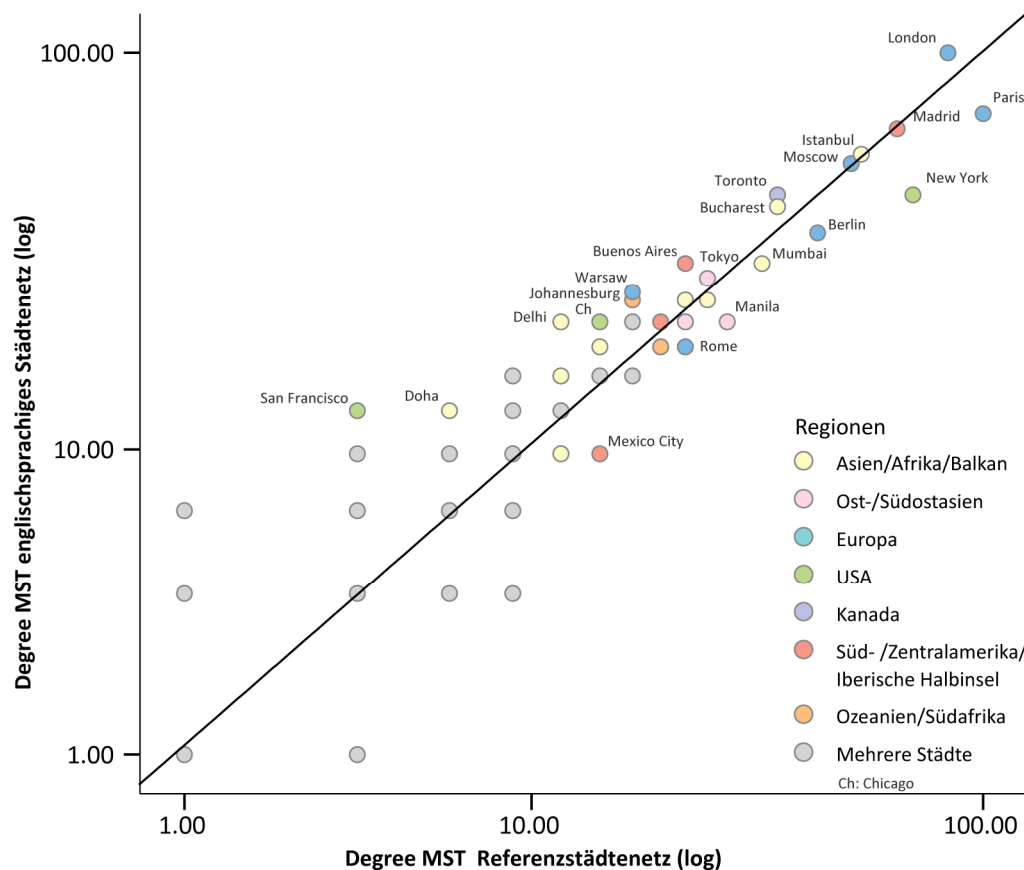
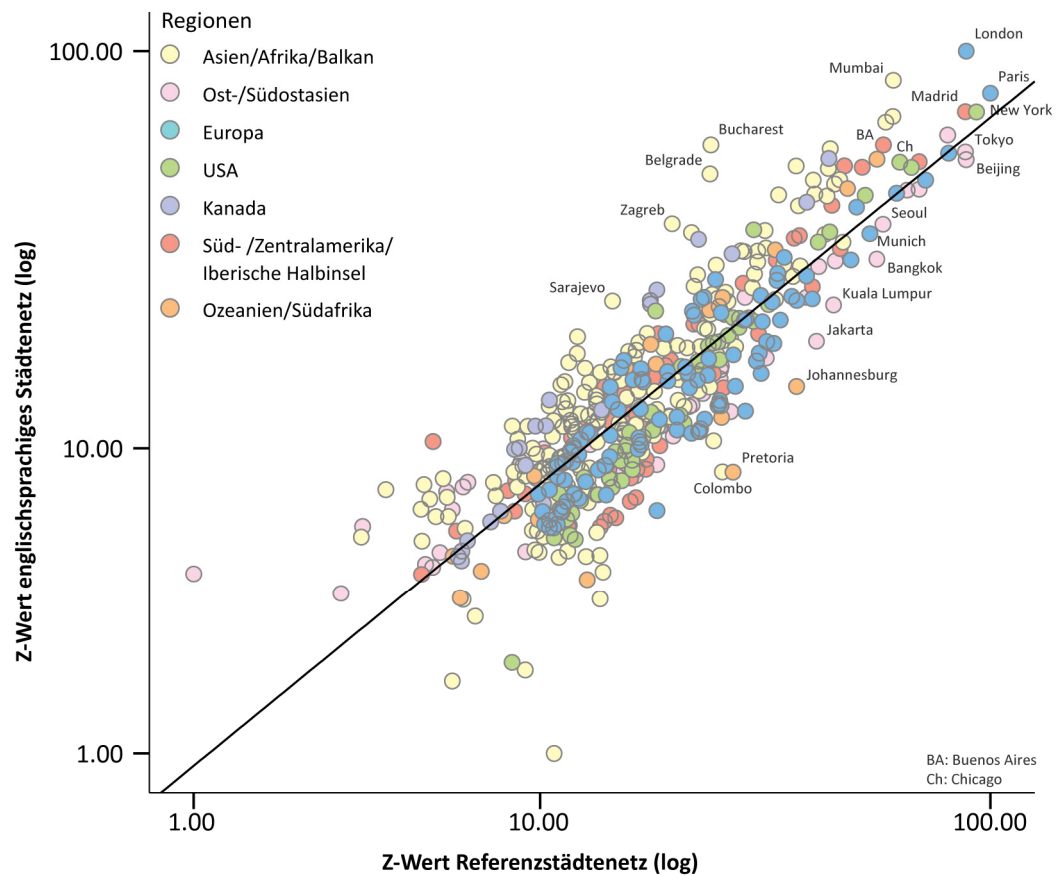


Abbildung 24: Scatterplot mit Regressionsgerade zwischen dem *degree* im MST des englischsprachigen und des Referenzstädtenetzes



Der letzte Vergleich basiert auf den Regionen, welche sich in den zwei Netzwerken herauskristallisieren. Regionen sind als Gruppen von engverflochtenen Städten definiert. In Abbildung 20 sieht man, dass das englischsprachige Städtenetz in sieben Regionen organisiert ist: USA, KANADA, SÜD-/ZENTRALAMERIKA/IBERISCHE HALBINSEL, EUROPA, ASIEN/AFRIKA/BALKAN, OST-/SÜDOSTASIEN und OZEANIEN/SÜDAFRIKA. Im Referenzdatensatz (Abbildung 21) lassen sich auf demselben Aggregationsniveau nur fünf Regionen identifizieren: NORDAMERIKA, SÜD-/ZENTRALAMERIKA/IBERISCHE HALBINSEL, EUROPA, ASIEN/AFRIKA und OZEANIEN.

Die systematische Gegenüberstellung der zwei regionalen Organisationen (Tabelle 18) zeigt wie viele Städte in beiden Netzwerken in die gleiche Region klassiert werden und welche Änderungen sich ergeben.

Tabelle 18: Vergleichsmatrix zwischen der regionalen Organisation im Referenz- und im englischsprachigen Städtenetz

Referenz Englischsprachig	Asien/Afrika	Europa	Nordamerika	Süd-/ Zentralamerika/ Iberische Halbinsel	Ozeanien
Asien/Afrika/Balkan	154	34			
Ost-/Südostasien	54				
Europa	1	107			
USA			52		
Kanada			21		
Süd/ Zentralamerika/ Iberische Halbinsel				72	
Ozeanien / Südafrika	5				15

Es sind drei Arten von Mustern erkennbar:

- 1 Stabilität: die Regionen sind in beiden Netzwerken gleich. Die SÜD-/ZENTRALAMERIKANISCH/IBERISCHE Region gehört zu dieser Kategorie.
- 2 Spaltung: eine Region im Referenznetzwerk teilt sich auf zwei Regionen im englischsprachigen Netzwerk auf. So teilt sich die NORDAMERIKANISCHE Region in eine KANADISCHE (21 Städte) und eine US-AMERIKANISCHE Region (52 Städte) auf. Die 54 OST-/SÜDOSTASIATISCHEN Städte spalten sich von der ASIATISCH-AFRIKANISCHEN Region ab, gleichermassen spalten sich die fünf südafrikanischen Städte ab und bilden mit den ozeanischen Städten die Region OZEANIEN/SÜDAFRIKA. Nach demselben Muster trennen sich die 34 balkanischen Städte von der EUROPÄISCHEN Region.
- 3 Fusion: ganze Regionen oder Teile des Referenznetzwerks fusionieren zu einer Region im englischsprachigen Netzwerk. So bilden die fünf von der ASIATISCH-AFRIKANISCHEN Region abgespaltenen südafrikanischen Städte eine gemeinsame Region OZEANIEN/SÜDAFRIKA mit 15 ozeanischen Städten. Ebenso vereinigen sich 34 Städte im Balkan mit 154 asiatischen und afrikanischen Städten zu einer Region ASIEN/AFRIKA/BALKAN.

Der Balkan und Südafrika können als Wechselregionen bezeichnet werden. Je nach Datengrundlage variiert ihre regionale Einbettung. Dieser von der Datengrundlage abhängige Unterschied ist jedoch mit der geographischen Lage dieser Regionen zu erklären: zwischen Europa und Asien für den Balkan, zwischen Afrika und Ozeanien für Südafrika.

Zusammenfassend ergibt der systematische Vergleich der netzwerkanalytischen Masse, dass die englischsprachige Wikipedia dem definierten Referenzdatensatz mehrheitlich entspricht. Die sprachlich bedingten räumlichen Verzerrungen sind für alle untersuchten Masse, mit Ausnahme des gewichteten *degrees* klein. Die englischsprachige Wikipedia ist somit hinsichtlich des sprachlichen Einflusses für die gewählten Methoden robust.

Basierend auf denselben netzwerkanalytischen Massen wird zudem die Robustheit des Datensatzes hinsichtlich drei weiterer Aspekte evaluiert (Tabelle 19):

- Volatilität der Einträge (Tabelle 19A): Es wird getestet wie sich die Masse verändern, wenn 25 % zufällig ausgewählte Links zwischen den Artikeln im betrachteten Datensatz gelöscht werden. Die Evaluation ergibt, dass die Datengrundlage auch in dieser Hinsicht robust ist.
- Modellierung der Stärke der Beziehung zwischen den Städten (Tabelle 19B): Es wird getestet wie sich die Masse verändern, wenn das Netzwerk aufgrund einer einfachen Normalisierung (siehe Formel 8) berechnet wird. Das Netzwerk ist auch in dieser Hinsicht robust.
- Anzahl betrachtete Städte (Tabelle 19C): Es wird getestet wie sich die Masse verändern, wenn das Netzwerk nur aus den 200 bzw. 1'000 zentralsten Städten besteht. Die Datengrundlage ist auch in dieser Hinsicht robust. Die grössten Abweichungen zeigen sich bei der regionalen Organisation: eine kleinere Anzahl Städte führt zu einer höheren Aggregation bzw. zu einer tieferen Auflösung. Städte, die im Netzwerk mit 1'000 Städten zu zwei getrennten, aber eng verflochtenen Regionen gehören, werden im Netzwerk mit 200 Städten zu einer Region aggregiert.

Tabelle 19: Evaluation der Robustheit des Datensatzes hinsichtlich A) der Volatilität der Einträge, B) der Modellierung der Stärken der Beziehungen zwischen den Städten und C) der Anzahl betrachteter Städte.

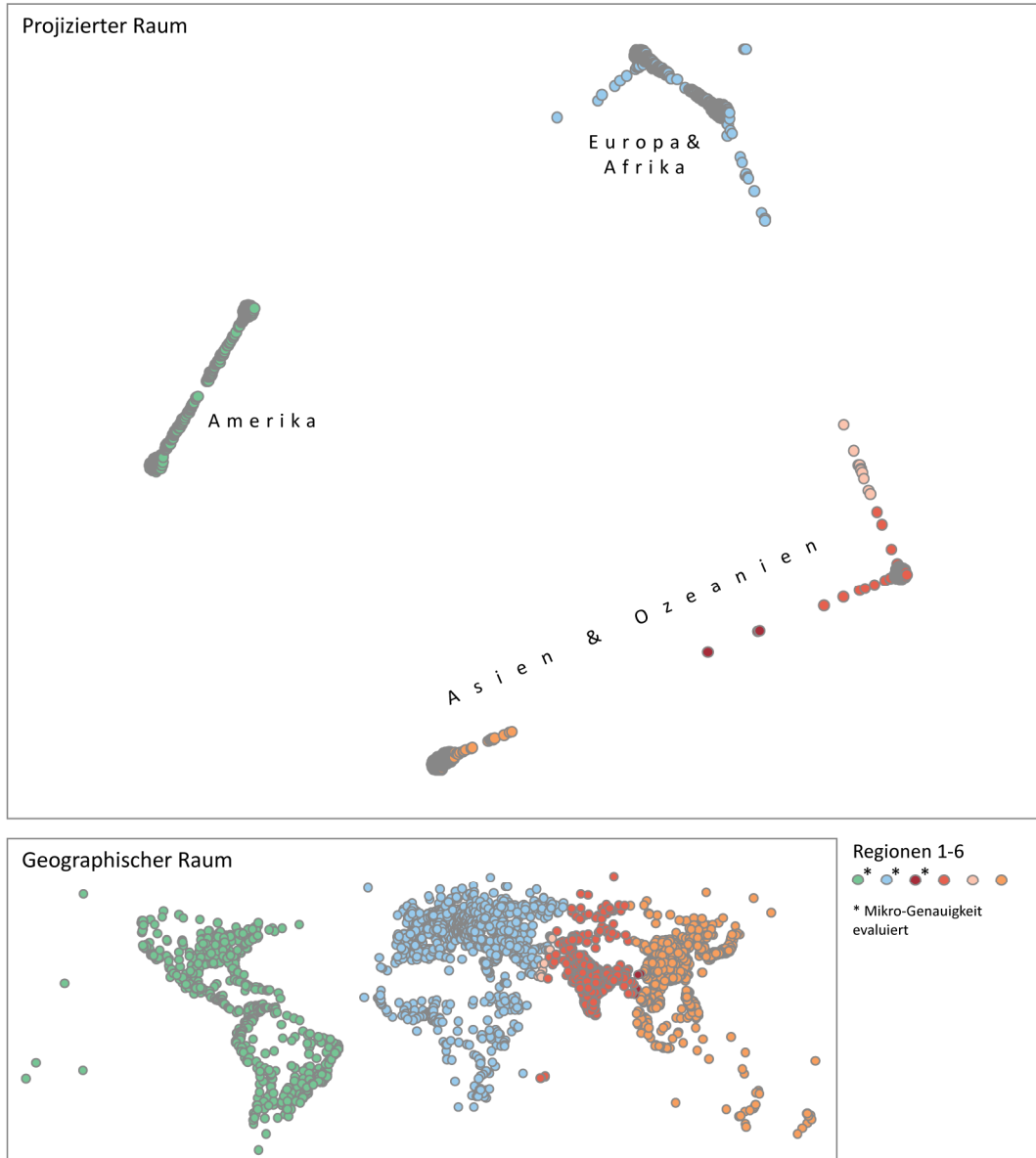
Datensätze	Korrelationskoeffizient nach Pearson			Hypergeom. Koeffizient
	Gewichteter <i>degree</i>	<i>Degree</i> im MST	Kantengewichte im MST	
A) Volatilität der Einträge				
Zufällige Auswahl an 75% der verbindenden Artikeln	1.000	0.990	0.931	0.000
B) Modellierung der Stärken der Beziehungen zwischen den Städten				
Einfache Normalisierung aufgrund der ausgehenden Links	0.998	0.994	0.914	0.000
C) Anzahl betrachteter Städte				
200 Städte	0.993	0.450	0.877	0.000
1'000 Städte	0.999	0.774	0.940	0.000

6.2 Verräumlichung mittels DrL/VxOrd-Algorithmus

Nachdem im vorherigen Kapitel die Robustheit der Datenbasis evaluiert wurde, widmet sich dieses Kapitel der Evaluation einer in der Arbeit oft verwendeten Methode: des DrL/VxOrd-Algorithmus. Dieser Algorithmus dient der Dimensionsreduktion und wird somit im Verräumlichungsprozess eingesetzt. Entscheidend für die Methodenevaluation ist, dass die mit der Verräumlichung hervorgehobenen Strukturen und erkannten Muster genau im zweidimensionalen Raum wiedergegeben werden. Ein topologischer zweidimensionaler Raum ist genau wiedergegeben, wenn die Lage der Objekte der erwarteten Lage entspricht. Wie in Kapitel 5.5 beschrieben wird insbesondere die Genauigkeit des zweidimensionalen Raumes in Abhängigkeit von der Anzahl Ähnlichkeitsbeziehungen, die für jedes Objekt in der Inputmatrix mit einbezogen werden, evaluiert. Nach Klavans und Boyak (2006a) sind die zweidimensionalen Räume am präzisesten, wenn die Inputmatrix für jedes Objekt die 15 stärksten Ähnlichkeitsbeziehungen enthält. In diesem Kapitel werden die Resultate der Evaluation dieser Eigenschaft der Inputmatrix vorgelegt. Zur Erinnerung: die Inputmatrix besteht aus 3'532 Städten und ihren messbaren Distanzen im geographischen Raum. Daraus werden drei topologische Räume mit dem DrL/VxOrd-Algorithmus generiert. Variiert wird dabei die Anzahl berücksichtigter Ähnlichkeitsbeziehungen pro Objekt. In diesem Fall entspricht dies der Anzahl der Nachbarstädte pro untersuchter Stadt. Der obere Teil der Abbildung 25 stellt den mit einer Inputmatrix von 500 Nachbarstädten pro Stadt generierten, topologischen Raum dar. Abbildung 27 (oberer Teil) wird mit 100 Nachbarstädten pro Stadt und Abbildung 28 mit 15 Nachbarstädten pro Stadt, wie von Klavans und Boyak (2006a) vorgeschlagen, generiert. Jedes Punktsymbol im topologischen Raum entspricht einer Stadt und die Farben entsprechen der regionalen Organisation (Meso-Ebene). Die unteren Teile der

jeweiligen Abbildungen stellen die im topologischen Raum identifizierten Regionen im geographischen Raum dar.

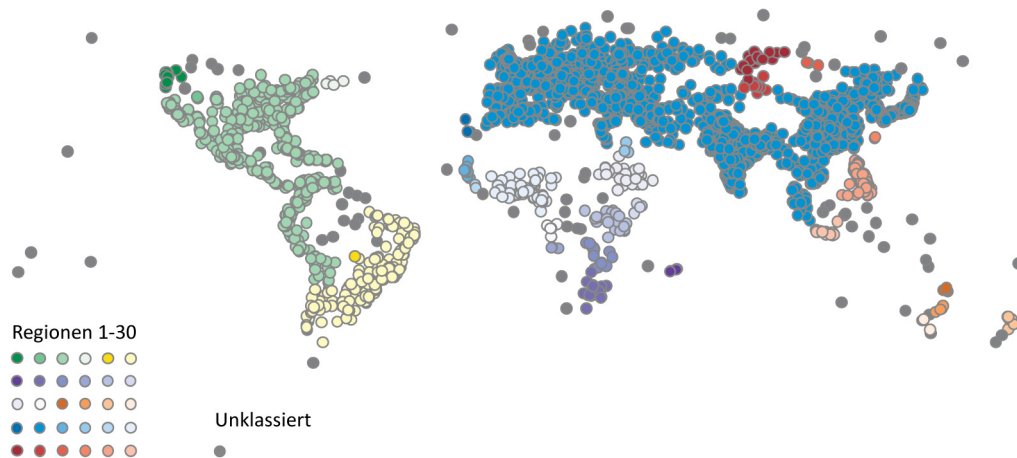
Abbildung 25: Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 500 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)



Die Projektion der Städte unter Berücksichtigung der 500 nächsten Nachbarstädte pro Stadt (Abbildung 25) zeigt ein Bild, in dem die Kontinentalstruktur erkennbar wird. Auf der Meso-Ebene werden sechs Regionen identifiziert: Amerika, Europa & Afrika, Zentralasien, Ostasien & Ozeanien und zwei sehr kleine Regionen. Der hypergeometrische Koeffizient beträgt 0.00 (siehe Kapitel 5.5). Daraus lässt sich schliessen, dass die im topologischen Raum identifizierten Regionen die Regionen im geographischen Referenzraum (Abbildung 26) reproduzieren. Trotzdem sind wichtige Unterschiede zwischen den zwei Meso-Ebenen festzustellen:

- 1 Im projizierten Raum (Abbildung 25) kristallisieren sich sechs Regionen heraus, während sich im Referenzraum (Abbildung 26) 30 Regionen bilden.
- 2 Die sechs amerikanischen Regionen in der Referenzdarstellung (Abbildung 26) werden zu einer amerikanischen Region im projizierten Raum (Abbildung 25) zusammengefügt.
- 3 Die euroasiatische Grossregion (Abbildung 26) teilt sich auf: Europa und Westasien gruppieren sich mit den 12 afrikanischen Regionen zu einer euroafrikanischen Region (Abbildung 25), der zentralasiatische Teilbereich bildet mit zwei kleineren Regionen die zentralasiatische Region und der ostasiatische Teilbereich, ergänzt durch acht kleinere Regionen aus Südostasien und Ozeanien, bildet die ostasiatische Region.

Abbildung 26: Städte im geographischen Raum mit Hervorhebung der Referenzregionen

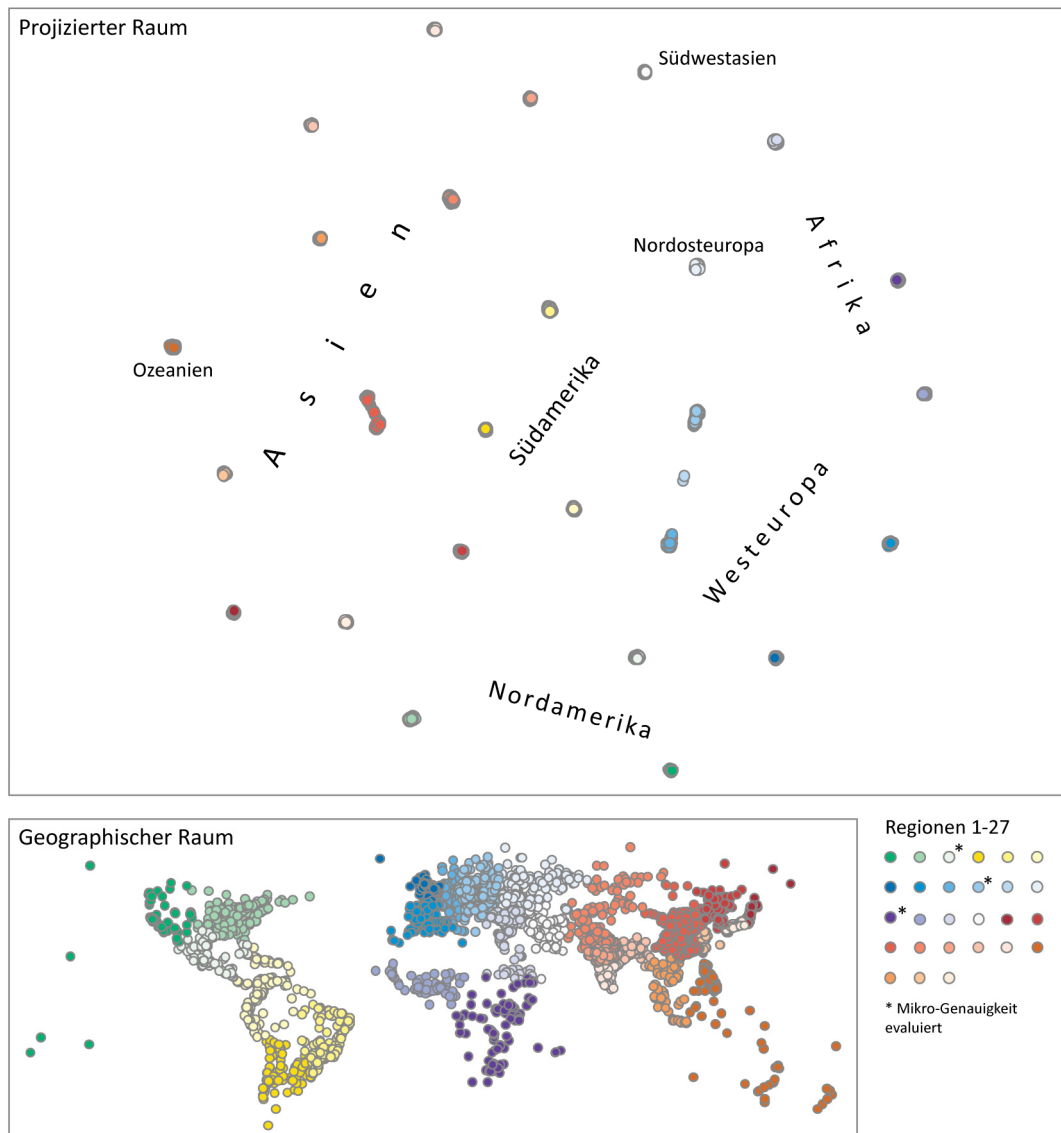


Die Übereinstimmung auf Makro-Ebene ist mit einem Korrelationskoeffizient von $r=0.638$, hoch. Die Distanzen zwischen den Regionen im topologischen Raum spiegeln die Distanzen im geographischen Raum wider. Amerika grenzt sich wie im geographischen Raum klar von Eurasien und Afrika ab. Grösser als im geographischen Raum sind die Distanzen zwischen der euroafrikanischen, der zentralasiatischen und der ostasiatisch-ozeanischen Region.

Auf Mikro-Ebene sind die Evaluationsergebnisse kontrastreicher. Die Übereinstimmung der intraregionalen Distanzen ist nur innerhalb der amerikanischen Region hoch ($r=0.856$). In der euroafrikanischen und insbesondere in der zentralasiatischen Region ist die Kongruenz zwischen den intraregionalen Distanzen im topologischen Raum und im geographischen Raum kleiner. Die Korrelationskoeffizienten betragen $r=0.530$ bzw. $r=0.332$.

Die Projektion der Städte unter Berücksichtigung der 100 nächsten Nachbarstädte zeigt sich in Abbildung 27.

Abbildung 27: Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 100 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)



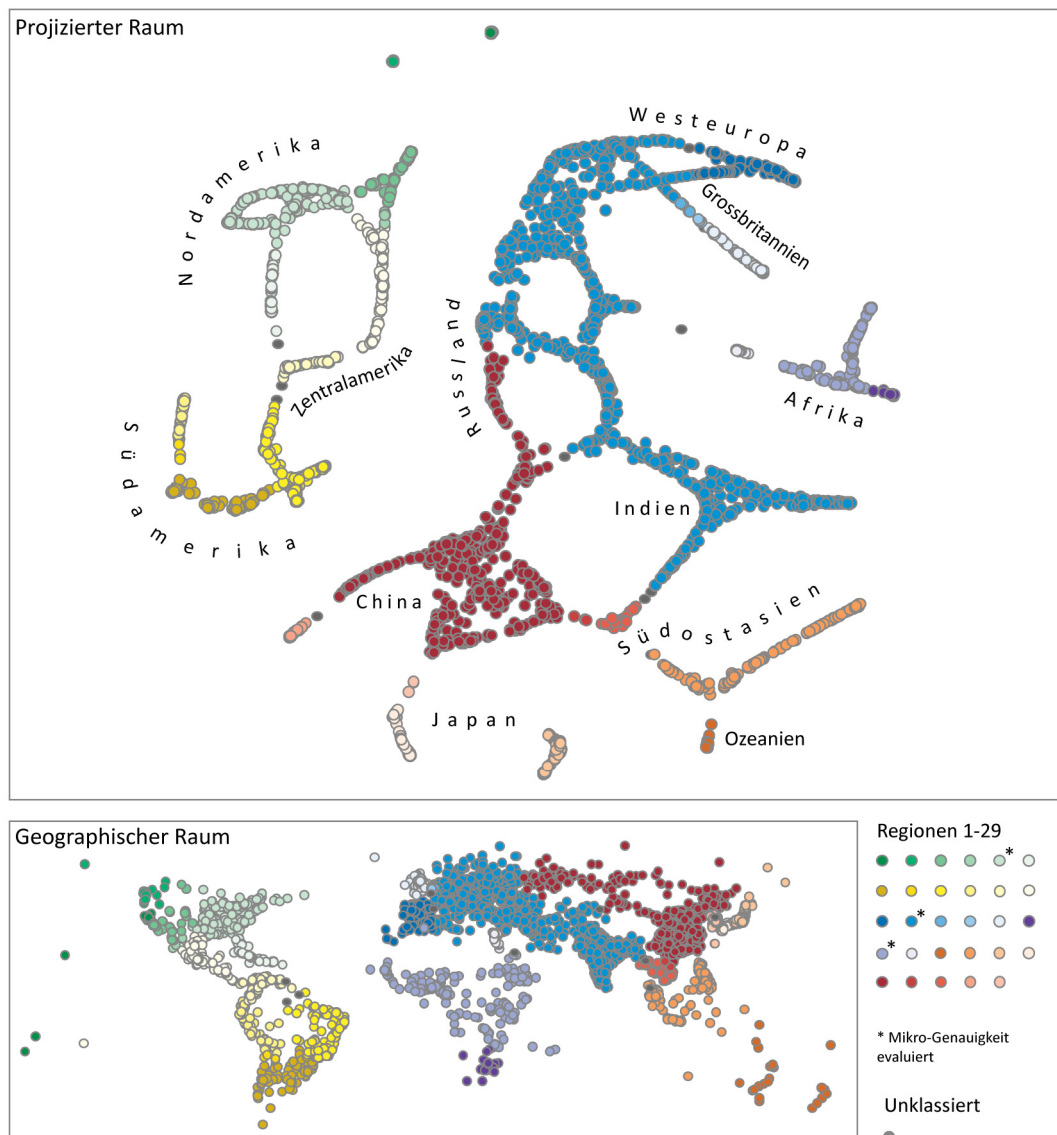
Die auffälligste Eigenschaft ist die Konzentration der 3'532 Städte auf 27 sehr dichte Regionen. Die Anzahl Regionen entspricht mit 27 fast genau der Anzahl Referenzregionen (30). Der hypergeometrische Koeffizient von 0.000 bestätigt wiederum, dass die Regionen im projizierten Raum die Referenzregionen (Abbildung 26) reproduzieren, trotzdem zeigen sich Unterschiede auf der Meso-Ebene. Einerseits teilen sich die euro-asiatischen und die panamerikanischen Referenzregionen (Abbildung 26) auf mehrere kleinere Regionen im projizierten Raum (Abbildung 27) auf, andererseits gruppieren sich die zwölf afrikanischen Regionen zu zwei Regionen und die sechs südostasiatischen und ozeanischen Regionen (Abbildung 26) zu einer Region im projizierten Raum (Abbildung 27).

Die Makrostruktur wird schwach wiedergegeben. Wie aus der Visualisierung ersichtlich ist, sind die Regionen dispers und gleichmässig innerhalb eines Kreises angeordnet. Die

Distanzen zwischen den Regionen sind tendenziell gleich und stimmen nur geringfügig mit den Distanzen im geographischen Raum überein. Dies wird mathematisch mit einem Korrelationskoeffizient von $r=0.296$ bestätigt. Eine Mikrostruktur ist auch nicht erkennbar. Jede Region wird auf einen Punkt reduziert, somit stimmen die intraregionalen Distanzen im topologischen Raum nur mässig mit den intraregionalen Distanzen im geographischen Raum überein. Dies widerspiegelt sich auch in den tiefen Korrelationskoeffizienten ($r=0.493$, $r=0.390$, $r=0.008$) zwischen den intraregionalen Distanzen in den zwei Räumen.

Im Gegensatz dazu präsentiert sich der Raum, welcher unter Berücksichtigung der 15 nächsten Nachbarstädte generiert wird, sehr differenziert und genau (Abbildung 28).

Abbildung 28: Verräumlichung der Städte aufgrund der gegenseitigen Luftliniendistanz unter Berücksichtigung der 15 nächsten Nachbarstädte pro Stadt (oben) und Hervorhebung der identifizierten Regionen im geographischen Raum (unten)



Auf der Meso-Ebene werden 29 Regionen identifiziert, eine weniger als in der Referenzdarstellung. Der hypergeometrische Koeffizient beträgt 0.000 und bestätigt wiederum, dass die Regionen im projizierten Raum (Abbildung 28) die Referenzregionen (Abbildung 26) reproduzieren. Trotzdem ändert sich auch in dieser Projektion die regionale Organisation. Die panamerikanische Referenzregion (Abbildung 26) teilt sich wiederum auf mehrere Regionen (Abbildung 28) auf, so wie sich die kleinen Referenzregionen in Afrika und Ozeanien (Abbildung 26) zu zwei bzw. einer Region gruppieren. Die Referenzregionen in Europa und Asien (exklusive Südostasien) (Abbildung 26) reorganisieren sich auf zwei grosse Regionen (Abbildung 28): Ostasien und Russland einerseits und Zentral- und Osteuropa und Westasien andererseits. Verschiedene Inseln (insbesondere Japan und Grossbritannien) und Halbinseln (insbesondere Korea und die Iberische Halbinsel) bilden eigenständige Regionen.

Auf der Makro-Ebene stimmen die interregionalen Distanzen im topologischen mit den Distanzen im geographischen Raum tendenziell ($r=0.594$) überein. Die Kontinentalstruktur ist klar erkennbar: Amerika (links), Eurasien (in der Mitte), Afrika (rechts) und Ozeanien (unten) sind klar voneinander getrennt. Gleichwohl ist die subkontinentale Struktur, insbesondere die Inseln klar erkennbar: Nordamerika unterscheidet sich von Zentral- und Südamerika, Japan, Korea und Südostasien von Asien, Grossbritannien von Europa.

Optimal wird insbesondere die Mikro-Ebene wiedergegeben. Die Korrelationskoeffizienten zwischen den intraregionalen Distanzen der drei untersuchten Regionen im projizierten und im Referenzraum betragen $r=0.983$ (Nordostamerika), $r=0.962$ (Europa & Südwestasien) bzw. $r=0.815$ (Afrika) und sind somit sehr hoch. Abbildung 29 hebt diese Zusammenhänge zwischen den intraregionalen Distanzen hervor. Die x-Achse stellt die intraregionalen Distanzen im geographischen und die y-Achse dieselben im projizierten Raum dar. Die Geraden entsprechen den drei Regressionsgeraden; eine für jede betrachtete Region. Aus den Regressionsgeraden zeigt sich, dass bei äquivalenter Zusammenhangsstärke (Korrelationskoeffizient zwischen $r=0.815$ und $r=0.962$) das Verhältnis zwischen den Distanzen im projizierten und im geographischen Raum je nach Region variiert. Die Steigung der Geraden drückt dieses Verhältnis aus. Das Verhältnis unterscheidet sich von den anderen Regionen stark für die betrachtete afrikanische Region. Während für die euroasiatische und die nordostamerikanische Region das Verhältnis zwischen geographischer und projizierter Distanz ca. 1:4 bzw. ca. 1:3 beträgt, entspricht dasselbe Verhältnis für die betrachtete afrikanische Region ca. 1:1. Dies bedeutet, dass die afrikanische Region im Verhältnis zu den anderen dichter projiziert wird. Die Erklärung dafür ist, dass durch die Projektion die Dichteverhältnisse nicht erhalten bleiben: wenig dichte Regionen im geographischen Raum (z.B. Afrika und Ozeanien) werden mit einer höheren Dichte projiziert.

Abbildung 29: Scatterplot mit Regressionsgeraden zwischen den intraregionalen Distanzen in drei Regionen im projizierten Raum (15 nächsten Nachbarn) und im Referenzraum

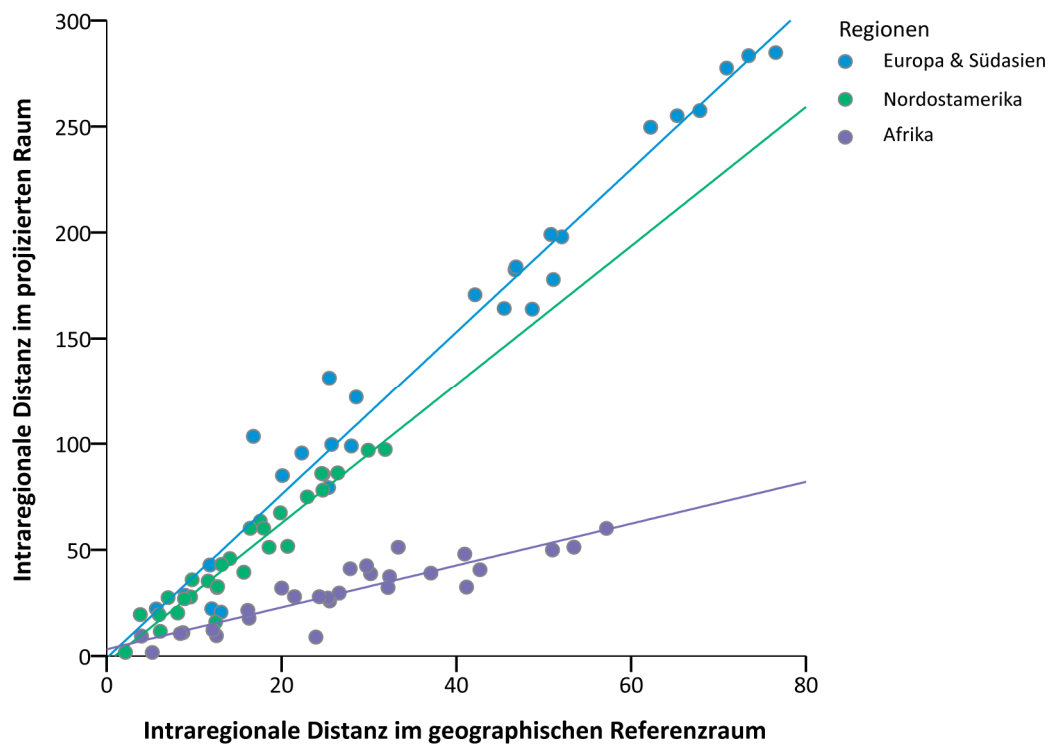


Tabelle 20 fasst die Genauigkeitsmasse für die drei projizierten Räume zusammen. Daraus lassen sich folgende Schlüsse ziehen:

- 1 Die Meso-Genauigkeit ist in den drei betrachteten Projektionen vergleichbar. Unterschiedlich ist jedoch die räumliche Auflösung bzw. die räumliche Granularität. Während mit 500 betrachteten nächsten Nachbarstädten sich nur sechs Regionen herauskristallisieren und somit die räumliche Granularität grob ist, zeigen die anderen Projektionen ca. 30 Regionen und somit eine höhere räumliche Auflösung.
- 2 Die Makro-Genauigkeit ist wohl für die Projektion mit 15 als auch mit 500 betrachteten nächsten Nachbarn hoch. Da, wie oben erwähnt, die Anzahl Regionen in der Projektion mit 500 viel tiefer ist, ergeben sich nur wenige interregionale Beziehungen. Die räumliche Granularität ist somit auch in diesem Fall größer für die Projektion mit 500 betrachteten nächsten Nachbarn.
- 3 Die Mikro-Genauigkeit ist für die Projektion mit 15 betrachteten nächsten Nachbarn am höchsten: sowohl kurze als auch lange intraregionale Distanzen werden sehr genau wiedergegeben.

Tabelle 20: Genauigkeitswerte für die drei betrachteten Projektionen

Anzahl nächste Nachbarn	Mikro-Genauigkeit *			Meso- Genauigkeit **	Makro- Genauigkeit *
	Region 1: Amerika	Region 2: Europa	Region 3: Afrika oder Asien		
15	0.983	0.962	0.815	0.000	0.594
100	0.330	0.493	0.008	0.000	0.296
500	0.856	0.530	0.332	0.000	0.638

* Pearson Korrelationskoeffizient, ** Hypergeometrischer Koeffizient

Das Postulat von Klavans und Boyack (2006a, 2006b) wird somit hinsichtlich der Genauigkeit der Verräumlichung bestätigt. Auch die Projektion eines anderen Datensatzes, welcher nur halb so viele Elemente beinhaltet, ist mit der Betrachtung von 15 nächsten Nachbarn am genauesten. Die Resultate zeigen gleichwohl, dass die Anzahl betrachteter nächster Nachbarn nicht nur die Genauigkeit sondern auch die Auflösung des projizierten Raumes bestimmt. Während die Projektion mit 15 nächsten Nachbarn eine sehr hohe Auflösung zeigt, werden mit 500 nächsten Nachbarn die Strukturen mit einer größeren räumlichen Granularität hervorgehoben. Die Meso- und Makro-Genauigkeiten der zwei Projektionen sind vergleichbar gut.

Die Projektion mit 100 nächsten Nachbarn ist am ungenauesten. Dies unterstreicht, dass der Zusammenhang zwischen Genauigkeit und der Anzahl betrachteter nächster Nachbarn nicht linear verläuft.

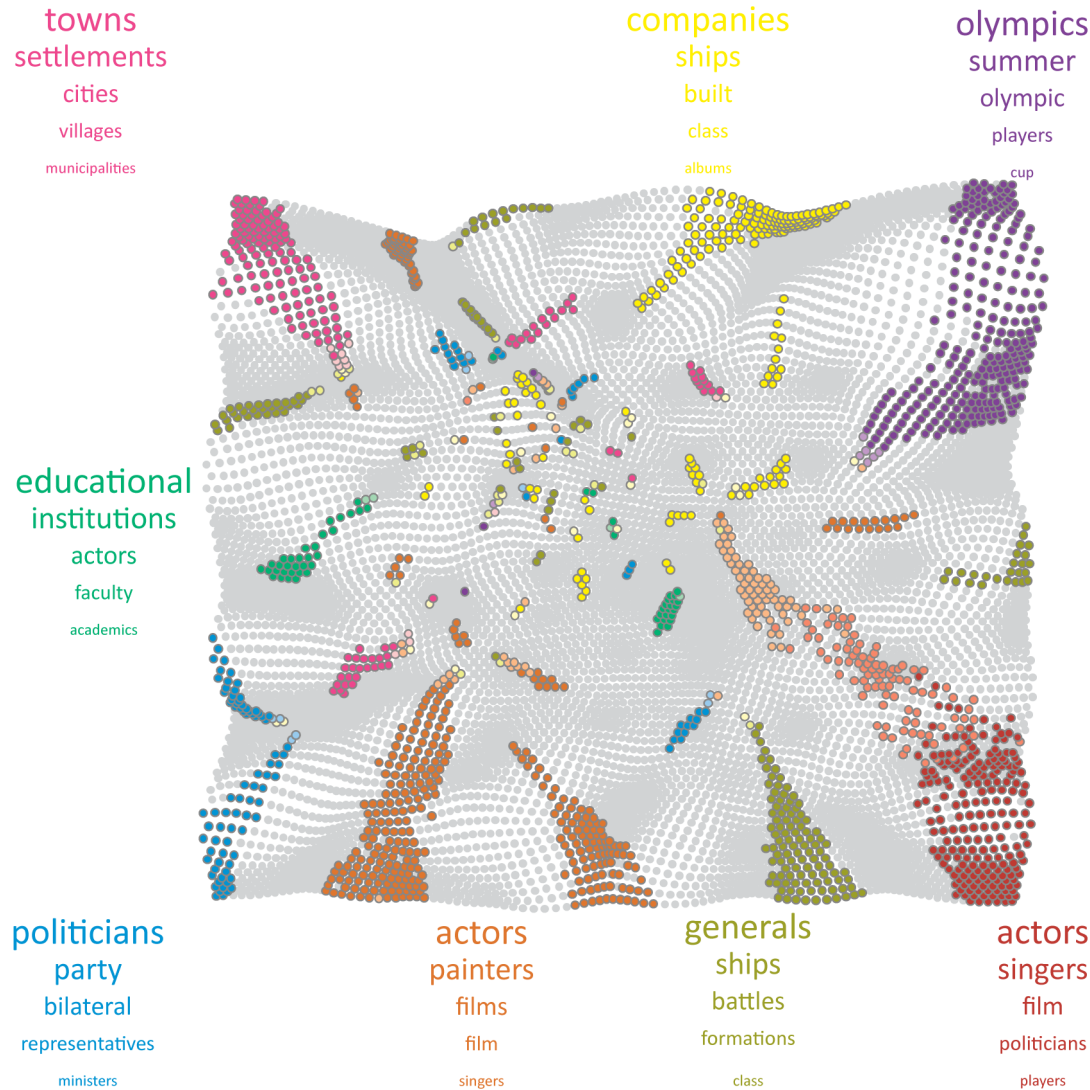
6.3 Definition der Funktionen

In den zwei vorhergehenden Unterkapiteln wird die englischsprachige Wikipedia als Datenbasis und die Genauigkeit der Verräumlichungen mit dem DrL/VxOrd-Algorithmus evaluiert. Dieses dritte Unterkapitel der Methodenevaluation widmet sich der Robustheit der Funktionen im Städtenetz. Funktionen sind als in sich thematisch homogene Gruppen definiert und umfassen Akteure, die zum selben Themenbereich gehören. Aufgrund der semantischen Ambiguität der Themen sind meistens unterschiedliche Lösungen und Abgrenzungen möglich. Im Folgenden werden zwei Definitionen der Funktionen einander gegenübergestellt, die sich auf zwei unterschiedliche Methoden stützen. Abbildung 30 zeigt die acht Funktionen, zu denen die Akteure aufgrund des Wikipedia-Kategoriensystems zusammengefasst werden. Der topologische Raum entspricht einer SOM (methodische Details siehe Kapitel 5.3). Dabei stellen die acht Farbtöne die acht Funktionen⁸ und die zwei Helligkeitsstufen die relative Häufigkeit der Funktion im Neuron (Punktsymbol) dar. Die Wortwolken stellen die fünf relevantesten (tf-idf) Begriffe pro Funktion dar (siehe Kapitel 5.3). Die grauen Punktsymbole stellen Neuronen dar, für die keine Akteure vorhanden sind.

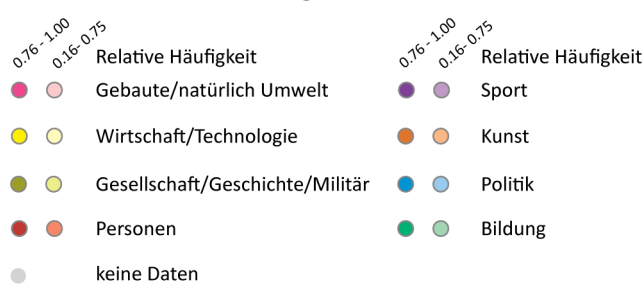
⁸ Die Funktionen und Subfunktionen werden im Text mit KAPITÄLCHEN geschrieben.

Die acht identifizierten Funktionen sind: GEBAUTE/NATÜRLICHE UMWELT, WIRTSCHAFT/TECHNOLOGIE, SPORT, PERSONEN, GESELLSCHAFT/GESCHICHTE/MILITÄR, KUNST, POLITIK und BILDUNG. Die Herleitung dieser Bezeichnungen sowie die inhaltliche Untersuchung der Funktionen folgt in Kapitel 7.2. Das Ziel dieses Kapitels ist hingegen das Eruiere der Robustheit der vorliegenden Klassierungslösung.

Abbildung 30: Verräumlichung der Funktionen mit der Themenbildung anhand des Wikipedia-Kategoriensystems mittels einer SOM



Funktionen und relative Häufigkeiten



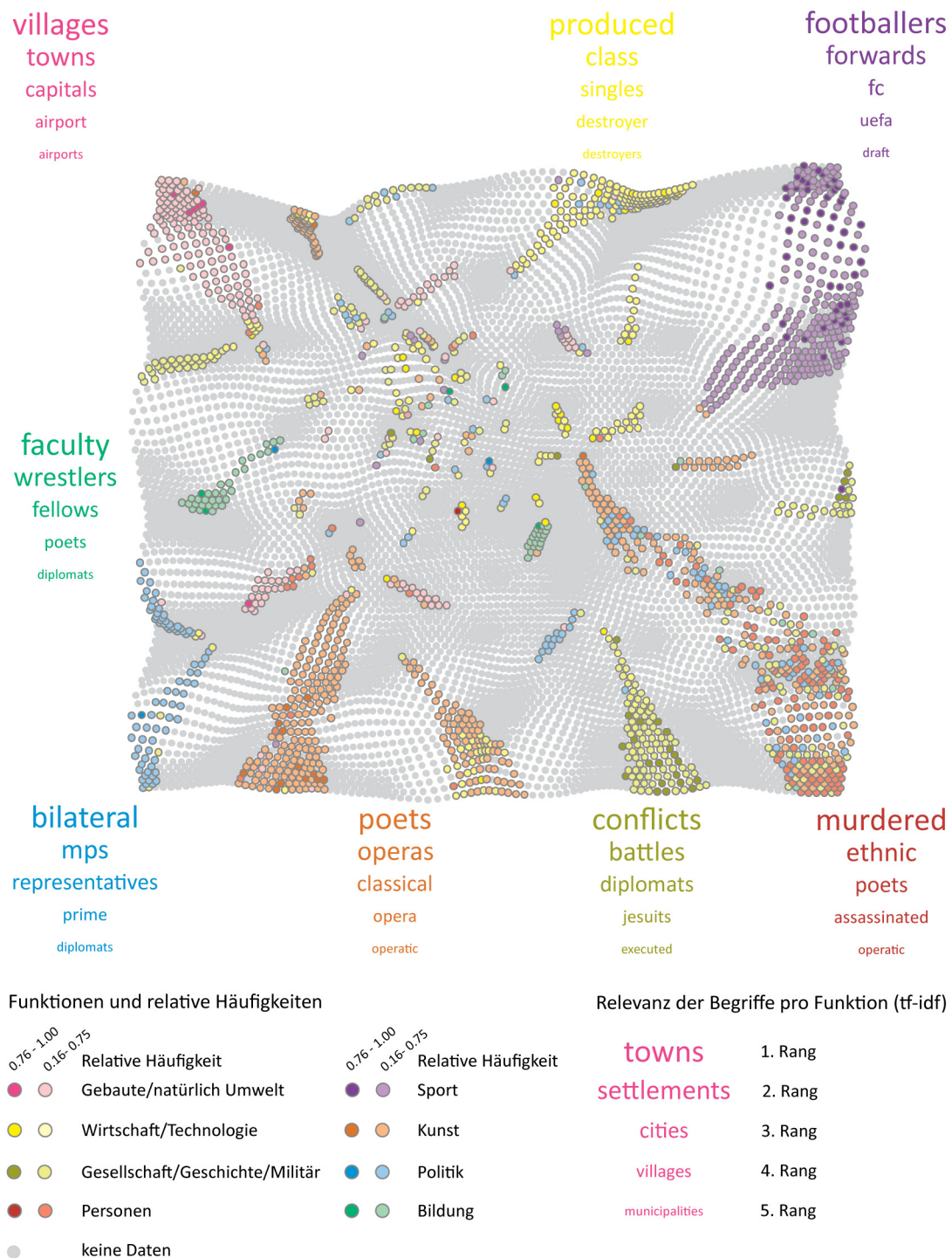
Relevanz der Begriffe pro Funktion (tf-idf)



Abbildung 31 zeigt die Funktionsdefinitionen, die sich auf eine Klassierung mittels *topic modelling* (siehe Kapitel 5.5) stützen, wobei die acht Funktionen auf den gleichen topolo-

gischen Raum wie in Abbildung 30 projiziert werden. Diese Abbildung dient als Vergleich.

Abbildung 31: Projektion der mit *topic modelling* definierten Funktionen auf die nach dem Wikipedia-Kategoriensystem berechnete Geometrie



Visuell stimmen die Muster in Abbildung 30 und Abbildung 31 grundsätzlich überein. Die meisten Neuronen gehören für beide Klassierungen derselben Funktion an. So gelangt z.B. die Region oben links (pink) zur Funktion GEBAUTE/NATÜRLICHE UMWELT, die Region oben rechts (violett) zur Funktion SPORT und die Regionen unten links

(blau) zur Funktion POLITIK. Unterschiedlich ist jedoch für die meisten Neuronen die relative Häufigkeit der Funktion. In Abbildung 30 liegt die Häufigkeit meistens über 0.75, während in Abbildung 31 die meisten Häufigkeiten geringer sind. Wenige Neuronen ändern zudem die funktionale Zugehörigkeit. Auffallend sind insbesondere die verschiedenen hellgrünen (GESELLSCHAFT/GESCHICHTE/MILITÄR) und blauen (POLITIK) Punktsymbole in der roten Region unten rechts in Abbildung 31. Nach der Funktionsdefinition mit Wikipedia werden diese als PERSONEN klassiert (rot in Abbildung 30). Die grundsätzliche Übereinstimmung wird sowohl durch den hypergeometrischen Koeffizienten als auch den Kappa Koeffizienten bestätigt. Der hypergeometrische Koeffizient beträgt 0.000 und bestätigt, dass die Funktionen nach *topic modelling* die Funktionen nach Wikipedia grundsätzlich reproduzieren. Der Kappa Koeffizient deutet mit einem Wert von 0.331 auf eine ausreichende Übereinstimmung der zwei Funktionsdefinitionen hin.

Tabelle 21 zeigt drei Messwerte, welche die Übereinstimmung der einzelnen Funktionen gemäss *topic modelling* mit dem Wikipedia-Kategoriensystem beschreiben. Daraus geht hervor, dass die Funktion SPORT in den beiden Klassierungen am besten übereinstimmt: der Anteil an Übereinstimmung (Formel 9) ist der grösste und der *error of omission* (nicht erkannte Elemente), sowie der *error of commission* (fälschlicherweise erkannte Elemente) sind die kleinsten verglichen mit den anderen Funktionen (für methodische Details siehe Kapitel 5.5). Der *error of commission*, d. h. das fälschliche Hinzufügen von fremden Elementen, zeigt sich auch im Vergleich zwischen Abbildung 30 und Abbildung 31. Konkret ersichtlich ist dieser Fehler an Punktsymbolen, die in Abbildung 31 violett sind, obwohl sie es in Abbildung 30 nicht sind. Damit werden Akteure hervorgehoben, die nach der *topic modelling*-Klassierung als SPORT (violett) klassiert werden obwohl sie nach der Wikipedia-Klassierung zu einer anderen Funktion gehören. Der *error of omission* zeigt sich an Punktsymbolen die in Abbildung 30 violett sind und in Abbildung 31 mit einem anderen Farbton oder mit einem helleren violett hervorgehoben werden. Dies sind Akteure, die nach der *topic modelling* Klassierung einer anderen Funktion zugehören, obwohl sie nach der Wikipedia-Klassierung als SPORT (violett) gelten.

Tabelle 21: Übereinstimmung der Funktionen nach Wikipedia (Referenzmodell) mit den Funktionen gemäss der *topic modelling*-Methode

Funktion	Anteil an Übereinstimmung	Error of omission	Error of commission
Sport	0.37	47.61	44.66
Wirtschaft/Technologie	0.32	51.49	52.10
Kunst	0.31	58.19	45.10
Gesellschaft/Geschichte/Militär	0.26	57.70	58.71
Gebaute/natürliche Umwelt	0.26	57.55	60.48
Politik	0.23	55.66	68.03
Bildung	0.18	55.37	76.92
Personen	0.14	78.18	69.87

Während für die Funktionen SPORT, WIRTSCHAFT/TECHNOLOGIE, sowie GESELLSCHAFT/GESCHICHTE/MILITÄR und GEBaute/NATÜRLICHE UMWELT der *error of omission* und der *error of commission* etwa gleich gross ist, zeigen die Funktionen POLITIK und BILDUNG einen höheren *error of commission*. Dies zeigt sich in Abbildung 31 insbesondere an den sehr vielen blauen Punktsymbolen in der mehrheitlich roten Region unten rechts. Dieser Wechsel entspricht Akteuren, die nach der Wikipedia-Klassierung als PERSONEN erkannt werden und nach der *topic modelling*-Klassierung zur Funktion POLITIK zählen. Der *error of commission* ist für die Funktion BILDUNG wenig ersichtlich, weil diese Funktion sehr wenig Akteure umfasst. Das gegenteilige Muster, ein grösserer *error of omission*, gilt für die Funktionen KUNST und PERSONEN. Ersichtlich ist der *error of omission* in Abbildung 31 an den gelben Punktsymbolen in der mehrheitlich orangen Region unten in der Mitte und an den blauen und hellgrünen Punktsymbolen in der mehrheitlich roten Region unten rechts. Diese entsprechen Akteuren, welche mit der *topic modelling*-Klassierung zur Funktion WIRTSCHAFT/TECHNOLOGIE (gelb) bzw. POLITIK (blau) oder GESELLSCHAFT/GESCHICHTE/MILITÄR (hellgrün) gezählt werden, obwohl sie in der Wikipedia-Klassierung zur Funktion KUNST (orange) bzw. PERSONEN (rot) gehören.

Die Übereinstimmung ist für die Funktion PERSONEN am geringsten. Viele Akteure werden mit der Klassierung nach *topic modelling* nicht als PERSONEN (rote Region unten rechts in Abbildung 31) erkannt (*error of omission* für PERSONEN), sondern als POLITIK (blaue Punkte), GESELLSCHAFT/GESCHICHTE/MILITÄR (hellgrüne Punkte) oder KUNST (orange Punkte) (*error of commission* dieser drei Funktionen), obwohl sie nach der Wikipedia-Klassierung zur Funktion PERSONEN (rote Punktsymbole in Abbildung 30) gezählt werden.

Eine zweite Ebene, auf der die zwei Funktionsdefinitionen verglichen werden können, ist der Inhalt der Artikel. Der Inhalt zeigt sich an den jeweils acht Wortwolken in Abbildung 30 und Abbildung 31. Dabei bestehen sie aus den jeweils fünf relevantesten Begriffen pro Funktion. Bei einem direkten Vergleich der Wortwolken sind die Ähnlichkeiten sehr gering. So stimmen für die Funktion GEBaute/NATÜRLICHE UMWELT die Begriffe *villages* und *towns* überein; die Begriffe *settlements*, *cities* und *municipalities* kommen jedoch nur in der Klassierung nach Wikipedia vor. Die Übereinstimmung ist für die weiteren Funktionen gleich oder noch geringer und sogar gleich null für die Funktionen SPORT, KUNST und PERSONEN. Der grosse Unterschied zwischen den Relevanzwerten (tf-idf) der zwei Klassierungen ist auch in Tabelle 22 ersichtlich. Die meisten nach der Wikipedia-Klassierung als relevanteste pro Funktion eingestuftten Begriffe haben einen tf-idf-Wert, der in der *topic modelling*-Klassierung Null entspricht. Der Grund dafür ist, dass bereits der Wechsel von wenigen Akteuren von einer Funktion zur anderen diesen Wert stark beeinflussen kann. Durch den Wechsel tendiert die *invers document frequency* (idf) gegen null und dadurch auch der tf-idf-Wert. Bestätigt wird diese Vermutung durch den Vergleich der absoluten Häufigkeitswerte (tf) (Tabelle 22). Die Häufigkeiten verändern sich trotz stark veränderter Relevanz für viele Begriffe nur leicht. Dies lässt sich wiederum am Beispiel der Funktion GEBaute/NATÜRLICHE UMWELT erläutern. *Settlements*, *cities*, *municipalities* haben eine sehr hohe Relevanz in der Klassierung nach Wikiped-

dia (12.51, 10.41 bzw. 5.15) und keine Relevanz in der *topic modelling*-Klassierung. Die Häufigkeit dieser drei Begriffe in der Funktion GEBaute/NATÜRLICHE UMWELT ändern sich jedoch nur geringfügig zwischen den zwei Klassierungen (*settlements*: von 1'476 zu 1'263, *cities*: von 2'562 zu 2'371, *municipalities*: von 608 zu 560). Grössere Änderungen in den Häufigkeiten sind bei einzelnen Begriffen ersichtlich: die Häufigkeiten für den Begriff *ships* in der Funktion GESELLSCHAFT/-GESCHICHTE/MILITÄR wechselt von 1'628 zu 522, analog dazu ändert sich die Häufigkeit von *educational* in der Funktion Bildung von 1'276 zu 870.

Tabelle 22: Vergleich der Relevanz- (tf-idf) und der absoluten Häufigkeitswerte relevanter Begriffe pro Funktion zwischen der Klassierung mit dem Wikipedia-Kategoriensystem und mit der *topic modelling*-Klassierung (TM)

Begriff	Funktion	Relevanz, tf-idf (Rang)		Häufigkeit, tf	
		Wikipedia	TM	Wikipedia	TM
generals	Gesellschaft/Geschichte/Militär	5.70 (1)	-	1'676	1'549
ships	Gesellschaft/Geschichte/Militär	5.54 (2)	-	1'628	522
battles	Gesellschaft/Geschichte/Militär	5.35 (3)	4.94 (2)	787	725
formations	Gesellschaft/Geschichte/Militär	4.21 (4)	-	875	627
class	Gesellschaft/Geschichte/Militär	4.16 (5)	1.04 (26)	1'224	329
actors	Personen	8.79 (1)	-	3'401	3'182
singers	Personen	4.66 (2)	-	2'657	2'990
film	Personen	3.05 (3)	-	2'846	1'864
politicians	Personen	2.67 (4)	-	1'522	384
players	Personen	2.48 (5)	-	679	410
towns	Gebaute/natürliche Umwelt	17.47 (1)	14.16 (2)	2'916	2'781
settlements	Gebaute/natürliche Umwelt	12.51 (2)	-	1'476	1'263
cities	Gebaute/natürliche Umwelt	10.41 (3)	-	2'562	2'371
villages	Gebaute/natürliche Umwelt	8.39 (4)	27.31 (1)	1'400	1'524
municipalities	Gebaute/natürliche Umwelt	5.15 (5)	-	608	560
actors	Kunst	9.32 (1)	-	4'293	3'087
painters	Kunst	7.00 (2)	-	3'226	2'831
films	Kunst	4.20 (3)	-	2'850	2'445
film	Kunst	3.97 (4)	-	4'409	4'433
singers	Kunst	3.61 (5)	-	2'454	1'279
politicians	Politik	8.87 (1)	-	2'782	2'767
party	Politik	6.30 (2)	-	1'975	1'923
bilateral	Politik	4.63 (3)	8.10 (1)	492	460
representatives	Politik	4.21 (4)	4.77 (3)	1'319	1'407
ministers	Politik	3.00 (5)	3.41 (7)	940	1'004
companies	Wirtschaft/Technologie	22.32 (1)	-	6'605	3'953
ships	Wirtschaft/Technologie	19.58 (2)	-	3'930	4'845

Begriff	Funktion	Relevanz, tf-idf (Rang)		Häufigkeit, tf	
		Wikipedia	TM	Wikipedia	TM
built	Wirtschaft/Technologie	15.19 (3)	-	3'049	3'690
class	Wirtschaft/Technologie	12.85 (4)	13.93 (2)	2'578	3'411
albums	Wirtschaft/Technologie	7.67 (5)	7.69 (7)	1'088	1'882
olympics	Sport	53.37 (1)	-	9'412	5'327
summer	Sport	52.23 (2)	-	9'210	5'260
olympic	Sport	49.46 (3)	-	8'722	5'191
players	Sport	28.17 (4)	-	7'021	5'092
cup	Sport	24.33 (5)	-	2'860	1'705
educational	Bildung	4.52 (1)	-	1'276	870
institutions	Bildung	4.27 (2)	-	1'207	788
actors	Bildung	3.92 (3)	-	459	592
faculty	Bildung	3.73 (4)	9.30 (1)	2'268	2'026
academics	Bildung	3.63 (5)	-	1'024	1'086

Neben der grundsätzlichen Übereinstimmung der Häufigkeiten deutet noch ein weiterer inhaltlicher Aspekt auf die Kongruenz der Funktionen: der thematische Kontext. Obwohl die Begriffe der Wortwolken nicht gleich sind, gehören sie meistens zum gleichen oder einem ähnlichen thematischen Kontext. So sind bei der Funktion GEBAUTE/NATÜRLICHE UMWELT nur zwei Begriffe gleich, jedoch stammen die weiteren Begriffe aus dem Kontext der Siedlungsformen (*settlements, cities, municipalities, capitals*) oder der Fluginfrastruktur (*airport, airports*). Für die Funktion SPORT stimmt kein Begriff überein, trotzdem gehören alle zehn Begriffe zu diesem Thema. Bei der Klassierung nach Wikipedia sind Begriffe aus dem Bereich Olympiade (*olympics, summer, olympic*) relevant und bei der Klassierung nach *topic modelling* aus dem Bereich Fussball (*footballers, fc, uefa*).

Sowohl der quantitative als auch der semantische Vergleich hat eine ausreichende Übereinstimmung der Funktion mit den zwei Klassierungsmethoden bestätigt. Die acht thematischen Schwerpunkte stimmen in beiden Klassierungen gut überein: GEBAUTE/NATÜRLICHE UMWELT, WIRTSCHAFT/TECHNOLOGIE, SPORT, PERSONEN, GESELLSCHAFT/GESCHICHTE/-MILITÄR, KUNST, sowie POLITIK und BILDUNG. Die genaue Abgrenzung dieser acht Funktionen ist jedoch je nach Klassierungsmethode unterschiedlich. Eine vollständige Übereinstimmung ist angesichts der Mehrdeutigkeit von semantischen Datensätzen auch kaum zu erwarten. Mehrere Lösungen sind möglich, ohne dass sie in Widerspruch zueinander stehen. Dies zeigt sich an einem einfachen Beispiel: *actors* und *poets* werden entweder zur Funktion KUNST oder zur Funktion PERSONEN gezählt. Welche der beiden Lösungen ist nun die Bessere? Wahrscheinlich sind beide je nach Kontext gleich angemessen, d.h. die Zuweisung zur einen oder anderen Funktion ist nicht falsch sondern nur unterschiedlich. Wichtig ist, dass diese Unter-

schiede konsistent sind. Aus diesem Grund werden im nächsten Absatz ausgewählte Unterschiede zwischen den beiden Klassierungen genauer analysiert.

Die relative Übereinstimmung ist für die Funktion PERSONEN am geringsten. Der *error of omission* und der *error of commission* sind hoch. Die Unterschiede sind in der roten Region unten rechts in Abbildung 30 und Abbildung 31 visuell ersichtlich. Drei Änderungsmuster sind dabei erkennbar:

1. Akteure, die nach Wikipedia-Klassierung zur Funktion PERSONEN gehören, werden nach der *topic modelling*-Klassierung zur Funktion POLITIK gezählt (rote Punktsymbole in Abbildung 30 werden zu blauen Punktsymbolen in Abbildung 31).
2. Akteure, die nach Wikipedia-Klassierung zur Funktion PERSONEN gehören, werden nach der *topic modelling*-Klassierung zur Funktion KUNST gezählt (rote Punktsymbole in Abbildung 30 werden zu orangen Punktsymbolen in Abbildung 31).
3. Akteure, die nach Wikipedia-Klassierung zur Funktion PERSONEN gehören, werden nach der *topic modelling*-Klassierung zur Funktion GESELLSCHAFT/-GESCHICHTE/MILITÄR gezählt (rote Punktsymbole in Abbildung 30 werden zu hellgrünen Punktsymbolen in Abbildung 31).

Erwartungsgemäss müssten sich diese wechselnden mehrdeutigen Akteure bei einer kohärenten Klassierung inhaltlich an der Schnittstelle zwischen PERSONEN einerseits und POLITIK, KUNST oder GESELLSCHAFT/-GESCHICHTE/MILITÄR anderseits befinden.

Die semantische Analyse der drei Wechselgruppen bestätigt diese Erwartung. Die relevantesten Begriffe (tf-idf) für diese drei Gruppen an Akteuren sind:

1. *princesses, diplomats, duchesses, archduchess, representatives* für die zu POLITIK wechselnden Artikel. Es handelt sich dabei um politisch aktive Personen.
2. *poets, classical, opera, superheroes, supervillains* für die zu KUNST wechselnden Artikel. Es handelt sich dabei mehrheitlich um Künstler oder fiktionale Charaktere.
3. *diplomats, executed, jesuits, priests, philosophers* für die zu GESELLSCHAFT/-GESCHICHTE/MILITÄR wechselnden Artikel. Die Übereinstimmung dieser Begriffe mit der Funktion GESELLSCHAFT/-GESCHICHTE/MILITÄR ist nicht direkt ersichtlich. Die Begriffe gehören vor allem dem religiös-philosophischen Themenbereich an. Die semantische Inkonsistenz besteht jedoch nur auf den ersten Blick. Für die Funktion GESELLSCHAFT/-GESCHICHTE/MILITÄR, definiert mit Wikipedia, sind neben militärisch-kriegerischen Aspekten, zu denen die fünf relevantesten Begriffe gehören auch religiöse Themen relevant. So gehören Begriffe wie *catholic* (tf-idf = Rang 8), *religious* (Rang 17) und *bishop* (Rang 18) zu den relevanten Begriffen der Funktion. Da die Funktion GESELLSCHAFT/-

GESCHICHTE/MILITÄR ein zweites Themenfeld, nämlich Religion umfasst, ist der Wechsel semantisch konsistent.

Zusammenfassend zeigt die Evaluation der Funktionsdefinitionen, dass die eruierten Hauptthemen nicht nur von der gewählten Methodik abhängen, sondern auch von anderen Methoden erkannt werden. Die Übereinstimmung der einzelnen Funktionen ist ausreichend. Vertiefere Analysen zeigen jedoch, dass die inhaltliche Klassierung von der semantischen Mehrdeutigkeit der Datengrundlage abhängt. Die acht identifizierten Funktionen sind somit eine mögliche und konsistente Lösung für die Definition der Funktionen im Städtenetz. Es ist nicht erstaunlich, dass vor allem die Funktion PERSONEN die geringste Übereinstimmung aufweist, weil sie generell inhaltlich unscharf ist und insbesondere mit SPORT (z.B. Sportler, Schiedsrichter), POLITIK (z.B. Minister, Botschafter), BILDUNG (z.B. Studenten, Alumni) und KUNST (z.B. Musiker, Schauspieler) grosse inhaltliche Überlappungen hat.

7 Resultate - Fallstudien

Nachdem im vorhergehenden Kapitel drei methodische Aspekte evaluiert worden sind, werden in diesem Kapitel die Resultate in der Form von drei Fallstudien dargelegt. Die erste Fallstudie untersucht das identifizierte Stdtenetz. Regionale Organisation, Zentralitt und Rollen der 3'532 Stdten werden analysiert. Die zweite Fallstudie befasst sich mit den Funktionen, zu denen die Akteure im Stdtenetz zusammengefgt werden knnen. Die Erkenntnisse aus Kapitel 6.3 werden aufgegriffen und bezogen auf die Forschungsfrage in inhaltlicher Perspektive vertieft. Die letzte Fallstudie fgt die Erkenntnisse aus den ersten zwei zusammen und hebt die Bedeutung der verschiedenen Funktionen fr die Hubs des Stdtenetzes hervor. Zustzlich werden die funktionalen Stdtenetze fr drei ausgewhlte Funktionen analysiert.

7.1 Das globale Stdtenetz

Regionale Perspektive

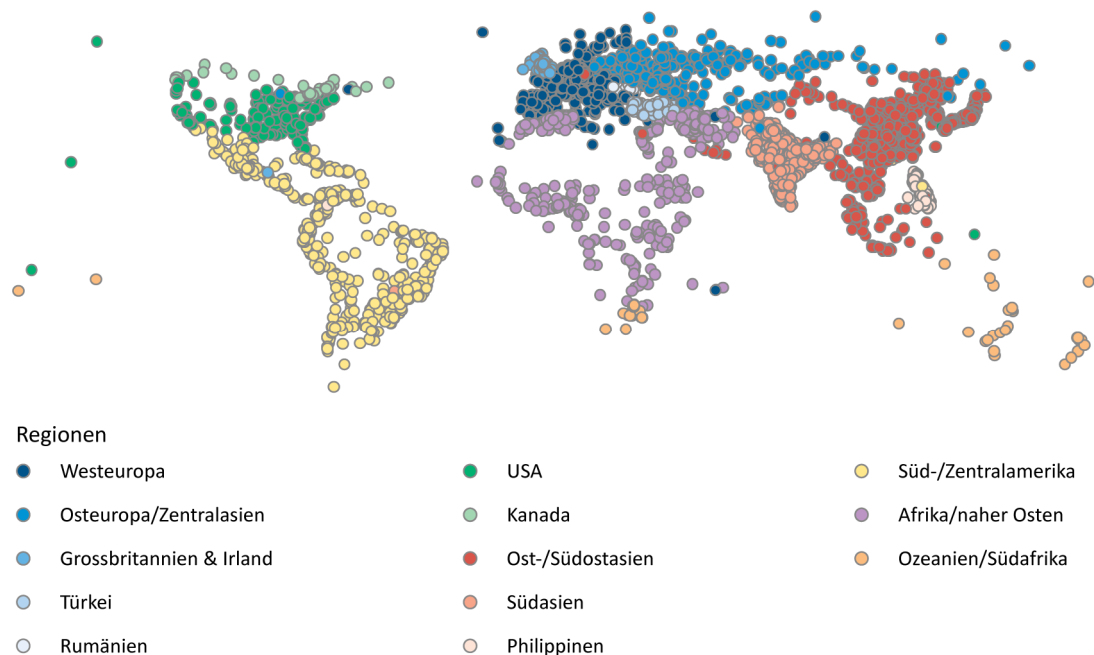
Analog zur geographischen Perspektive, in der Stdte zu Regionen gehren, organisieren sich die Stdte auch aus relationaler Perspektive zu kleineren und grsseren Regionen. Die 3'532 Stdte bilden 13 Regionen⁹, wobei die grsste 842 und die kleinste 25 Stdte umfasst. Wie in den nchsten Abschnitten kurz beschrieben wird, teilen sich die dreizehn Regionen in Subregionen auf. Im weiteren Verlauf der Arbeit beziehen sich die Resultate und die Diskussion jedoch ausschliesslich auf die dreizehn Regionen der

⁹ Die relationalen Regionen werden im Text mit KAPITLCHEN geschrieben.

höchsten Aggregationsstufe, obwohl eine umfassende Untersuchung des globalen Städtetetzes mehrere Ebenen gleichzeitig berücksichtigen sollte. Die inhaltliche Einschränkung beruht auf der methodischen Ausrichtung der Arbeit. Der gewählte Ansatz lässt sich jedoch in weiterführenden inhaltlichen Untersuchungen auf die mehr-Ebenen Analyse übertragen.

Bei einer Projektion der relationalen Regionen auf eine Weltkarte (Abbildung 32) wird die hohe geographische Kohärenz ersichtlich: Städte der gleichen Region sind auch geographisch benachbart und die Regionsgrenzen folgen meist den Staats- und Kontinentengrenzen.

Abbildung 32: Projektion der relationalen Regionen im geographischen Raum



Kleine geographische Inkonsistenzen, welche einzelne periphere Städte mit tiefer Zentralität betreffen, sind weniger das Resultat von besonderen Beziehungen, sondern eher die Konsequenz von Mängeln in den Daten. Die zwei Hauptgründe dafür sind das erhöhte Fehlerisiko bei Artikeln mit wenigen einkommenden Links und die Ambiguität gewisser Toponyme (z.B. Cartagena in Spanien, Kolumbien oder Chile), was zu falschen Links in Wikipedia führen kann. Abgesehen von diesen einzelnen Unsicherheiten ergeben sich kohärente und teilweise unerwartete Strukturen.

In Nordamerika folgen die zwei Regionen der Landesgrenze und enthalten einerseits die US-AMERIKANISCHEN und andererseits die KANADISCHEN Städte. ZENTRAL-/SÜDAMERIKA bilden eine Region, welche sich jedoch auf einem tieferen Organisationsniveau in drei Subregionen organisiert: 1) Südamerika, 2) Zentralamerika und 3) Mexiko und Karibik.

Der ozeanische Kontinent bildet mit den südafrikanischen Städten eine eigene Region, die in zwei Subregionen unterteilt ist (Südafrika und Ozeanien inkl. Inseln). Die ost- und

südostasiatischen Städte sind mit Ausnahme der Philippinen in einer Region organisiert. Zur selben Region gehören auch die Städte der kleinen Anrainerstaaten des persischen Golfes (Kuwait, Saudi-Arabien, Bahrain, Katar, Vereinigte Arabische Emirate und Oman). Auf subregionaler Ebene organisieren sich die koreanischen und die japanischen Städte zu zwei getrennten Subregionen. Besonders bemerkenswert ist die Stellung der Stadt Lhasa (Tibet), die nicht Teil der OST-/SÜDOSTASIATISCHEN, sondern der WESTEUROPÄISCHEN Region ist. Dies ist vermutlich mit der einmaligen politischen Situation in Tibet zu erklären. Die SÜDASIATISCHE Region ist auf subregionaler Ebene besonders fragmentiert und teilt sich auf fünf Subregionen auf. Die grösste Subregion umfasst die indischen und die sri-lankischen Städte, mit zwei Ausnahmen: die Städte des Bezirks Nadia (Westbengalen, Indien) bilden eine eigene Subregion und die Städte in der Region Punjab befinden sich in einer gemeinsamen Subregion zusammen mit den pakistanischen Städten. Die Gründe sind vermutlich historisch-politischer Natur. Die afghanischen und die bengalischen Städte bilden zwei eigenständige Subregionen.

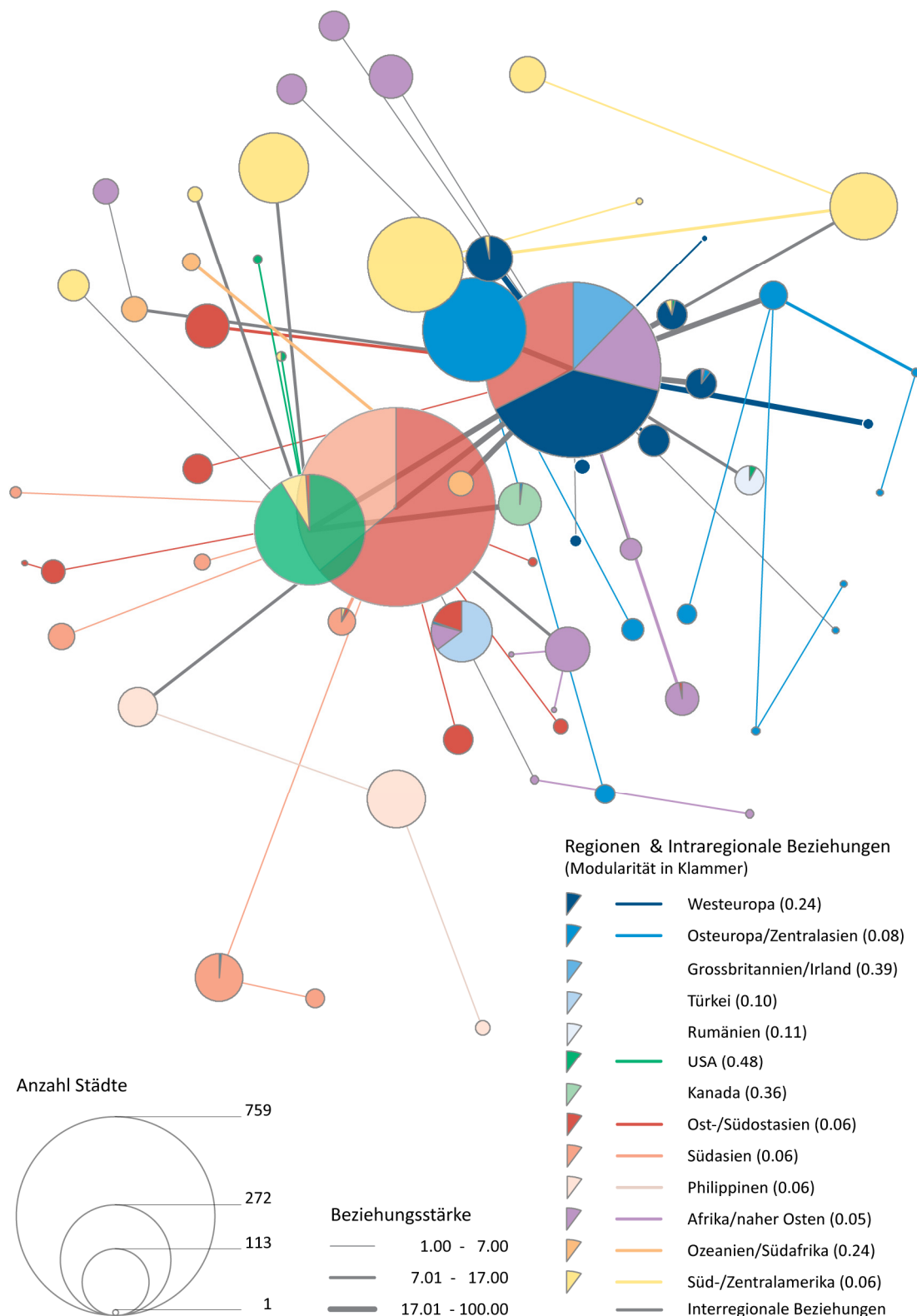
Die OSTEUEPÄISCH/ZENTRALASIATISCHE Region besteht aus den zwei Subregionen: Polen und ehemalige UdSSR. Auf dem europäischen Kontinent erstrecken sich neben dem westlichen Teil der OSTEUEPÄISCH/ZENTRALASIATISCHEN Region vier weitere Regionen: 1) GROSSBRITANNIEN/IRLAND, 2) TÜRKEI 3) RUMÄNIEN und 4) die WESTEUROPÄISCHE Region. Die WESTEUROPÄISCHE Region umfasst neben den Städten auf dem europäischen Kontinent auch die tunesischen und libyschen Städte, sowie die Städte der Réunion (französisches Überseedepartement). Die Gründe sind vermutlich die koloniale Vergangenheit und die administrativen französischen Gegebenheiten.

Die AFRIKANISCH/NAHÖSTLICHE Region umfasst den afrikanischen Kontinent mit Ausnahme der südafrikanischen, libyschen und tunesischen Städte, einen Grossteil der arabischen Halbinsel, Irak und Iran. Diese grosse geographische Ausdehnung geht Hand in Hand mit einer hohen inneren Fragmentierung. Acht Subregionen mit komplexen und unscharfen Grenzen bilden diese Region. Die grösste Subregion umfasst die Städte der Sub-Sahara, mit Ausnahme der nigrischen Städte (eigenständige Subregion). In Nordafrika sind die marokkanischen und die algerischen Städte in zwei Subregionen organisiert. Vergleichbar ist die Lage im Osten: die irakischen und armenischen einerseits und die iranischen Städte andererseits bilden zwei Subregionen. Israels Städte bilden eine eigenständige Subregion. Die restlichen Städte, alle aus dem westarabischen Teil der Region (Ägypten, Sudan, Somalia, Saudi-Arabien, Syrien, Libanon) bilden eine weitere Subregion.

Die Projektion derselben Städte in einen topologischen Raum (Abbildung 33) ist wiederum als eine Weltkarte zu verstehen, auf der die Lage und die regionale Organisation der Städte die Struktur dieser relationalen Welt zur Erscheinung bringen. Die Farben entsprechen den Regionen. Die Kreisdiagramme, Pole genannt, stellen Gruppen sehr eng verknüpfter Städte dar und die Liniensymbole den MST zwischen den Städten in den einzelnen Polen. Ein Lesebeispiel: Der grössere, mehrheitlich grüne Pol (in der Mitte links) umfasst 272 Städte in den USA (grün), sowie ZENTRAL-/SÜDAMERIKA (gelb).

Auf dem kleineren, hellgrünen Pol ballen sich die KANADISCHEN Städte. Dieser hat die Hauptinteraktionsachse zum oben erwähnten amerikanischen Pol. Die Lage eines Pols in diesem relationalen Raum wird hingegen basierend auf allen Interaktionen bestimmt. Die Zwischenlage der KANADISCHEN Städte, zwischen den drei grossen Polen suggeriert somit, dass die KANADISCHEN Städte zusätzlich zu den USA auch einen intensiven Austausch mit WESTEUROPA (dunkelblau), OST-/SÜDOSTASIEN (rot) haben.

Abbildung 33: Regionen des globalen Stdtenetzes im topologischen Raum



Das Zentrum der relationalen Welt besteht aus drei Hauptpolen (grosse zentrale Kreise), die knapp die Hlfte der Stdte umfasst.

Zum ersten (von links nach rechts, in grn) gehren 272 Stdte aus den USA und der Karibik. Dieser Pol positioniert sich sehr nahe am zweiten; dieser zweite schliesst die

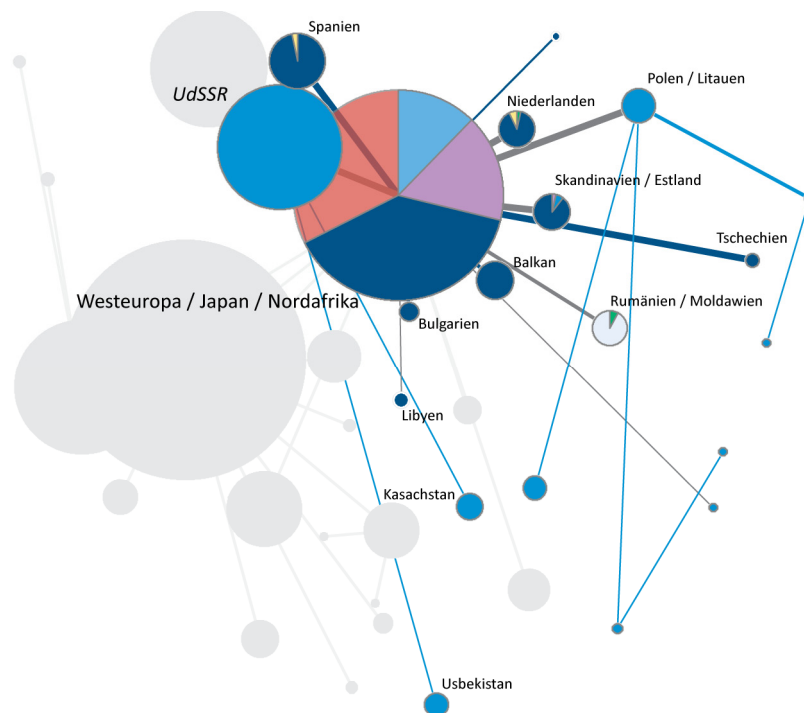
südostasiatischen und eine Mehrheit der chinesischen und indischen Städte (Total 759) ein. Trotz der Nähe im topologischen Raum zueinander, was ein Zeichen für hohe Interaktionsintensität ist, pflegen beide Pole die stärkste Beziehung (Hauptbeziehung) zum dritten Pol. Dieser umfasst eine Mehrheit der WESTEUROPÄISCHEN Städte, der japanischen und der nordafrikanischen Städte sowie alle Städte der Region GROSSBRITANNIEN/IRLAND (Total 608). Die zwei Hauptbeziehungen gehören zu den stärksten im System und bauen sich zwischen Städten unterschiedlicher Regionen auf.

Die weiteren Städte verteilen sich auf 60 Pole, die sich in abnehmender Zentralität um die drei Hauptpole organisieren und mit einem davon direkt oder indirekt verbunden sind, wobei die Mehrheit eine Hauptverbindung zum Europa-zentrischen Pol hat.

Bei genauerer Betrachtung der Regionen stellt man fest, dass sich die USA, mit Ausnahme von wenigen Ausreißern aus Oklahoma und Texas, auf einen Pol konzentrieren und somit eine sehr kompakte Region bilden. Gleich ist die Situation für KANADA. Die normalisierten Modularitätswerte (Formel 4) bestätigen die hohe Kohäsion dieser zwei Regionen (Abbildung 33).

Die WESTEUROPÄISCHEN Städte (Abbildung 34) positionieren sich zentral in und um den Europa-zentrischen Hauptpol. Auf der rechten Seite ist das Gebiet durch eine hohe Polyzentralität geprägt: Skandinavien, Osteuropa und der Balkan verteilen sich auf mehrere kleine Pole mit mehrheitlich starken Hauptbeziehungen zum Hauptpol. Polen splittet sich auf mehrere vor allem periphere und kleine Pole auf. Die Polyzentralität spiegelt sich in der durchschnittlich-hohen Modularität (0.24) wider.

Abbildung 34: Regionen des globalen Städtetetzes im topologischen Raum: Ausschnitt Europa und Zentralasien

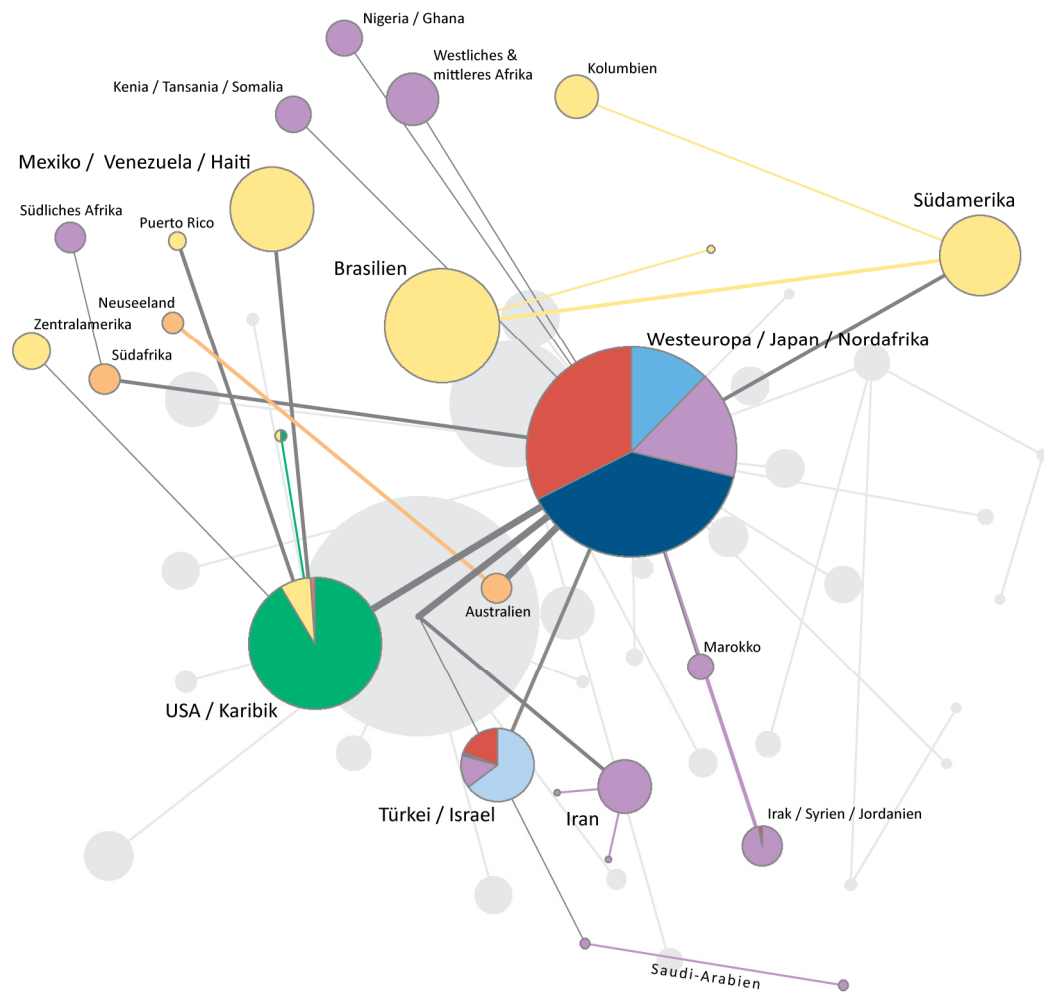


Die Verteilung der OSTEUPÄISCH/ZENTRALASIATISCHEN Region im topologischen Raum (Abbildung 34) offenbart die Dreiteilung dieser Region:

- Polen, Litauen und der Westen der Ukraine sind im topologischen Raum Teil des osteuropäischen Gebiets und haben starke Beziehungen zum Europa-zentrischen Hauptpol.
- Kasachstan und Usbekistan positionieren sich in einem peripheren Übergangsgebiet zwischen Europa und Asien und pflegen schwache Beziehungen zum russischen Pol.
- Russland und die angrenzenden Staaten konzentrieren sich auf einen grossen Seitenpol mit schwachen Beziehungen zum Europa-zentrischen Hauptpol.

Die aufgrund der sehr tiefen Modularität zu vermutende Fragmentierung zeigt die Region AFRIKA/NAHER OSTEN auch im topologischen Raum (Abbildung 35). Nicht nur verteilen sich die Städte auf neun Pole, sondern diese befinden sich zudem in unterschiedlichen Gebieten: die arabischen Städte (ohne die kleinen Anrainerstaaten des persischen Golfes) positionieren sich sowohl im Europa-zentrischen Hauptpol als auch peripher im Übergangsgebiet zwischen Europa und Asien; die Städte der Sub-Sahara befinden sich in peripherer Lage analog zu den SÜD-/ZENTRALAMERIKANISCHEN Städten.

Abbildung 35: Regionen des globalen Städtenetzes im topologischen Raum: Ausschnitt südliche Welthalbku­gel



Die arabischen Städte sind in sehr heterogene und komplexe Strukturen eingebaut. Im Hauptpol befinden sich die Städte aus Ägypten, Sudan, Jemen, den arabischen Emiraten und Tunesien (letztere sind Teil der westeuropäischen Region). Zum Europa-zentrischen Hauptpol beziehen sich die marokkanischen, irakischen, syrischen, jordanischen und libyschen Städte. Iran und Saudi-Arabien befinden sich im selben Übergangsgebiet, beziehen sich jedoch auf den Asien-zentrischen Hauptpol. Im Asien-zentrischen Hauptpol befinden sich die Städte der kleinen Anrainerstaaten des persischen Golfes, welche Teil der OST-/SÜDOSTASIATISCHEN Region sind. Israel und die Türkei bilden einen gemeinsamen Pol, der sich im asiatisch-nordamerikanischen Einflussgebiet befindet, was auf enge Interaktionen hindeutet. Die Hauptbeziehung pflegt dieser Pol jedoch zum Europa-zentrischen Hauptpol.

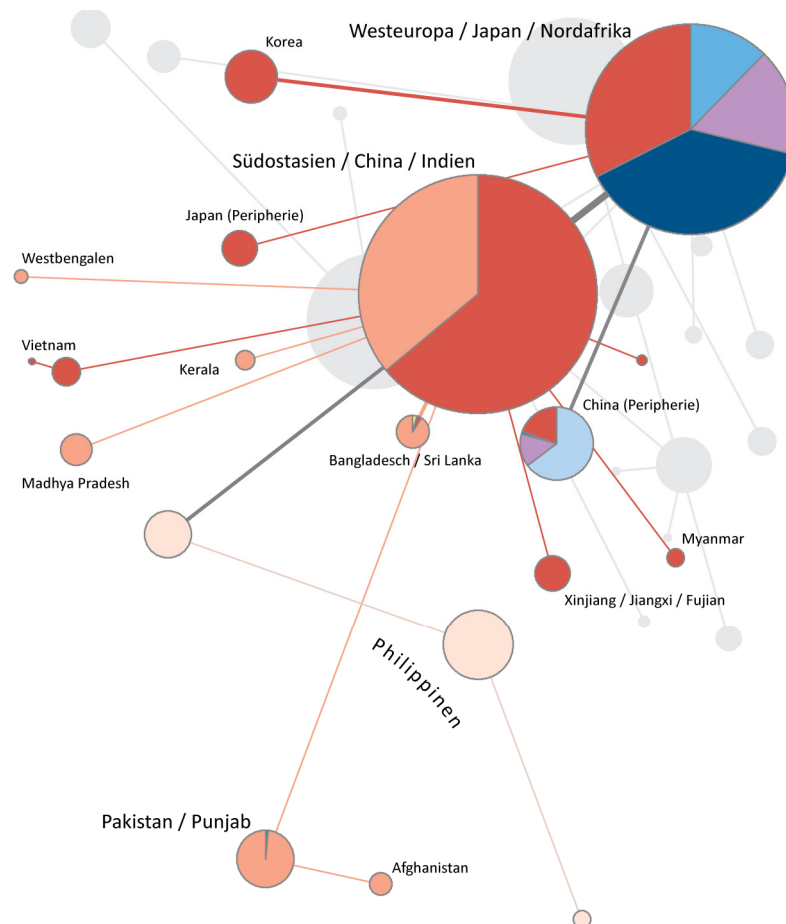
Die SÜD-/ZENTRALAMERIKANISCHEN Städte umspannen mit den Städten der Sub-Sahara und den neuseeländischen Städten ein grosses Gebiet, welches vom US-zentrischen Hauptpol bis zur Peripherie des Europa-zentrischen Hauptpols geht. Zentralamerika, die Karibik und Mexiko beziehen sich auf den US-zentrischen Hauptpol; Kuba, Jamaika und Panama sind sogar ein integraler Bestandteil davon. Südamerika orientiert sich stärker am Europa-zentrischen Hauptpol. Brasilien ist, trotz der grossen

Nähe zum Europa-zentrischen Hauptpol und insbesondere zu Spanien, jedoch nur indirekt damit verbunden. Kolumbien ist der periphere Pol.

Südafrika, Neuseeland und Australien sind zwar Teil derselben Region, teilen sich jedoch auf drei unterschiedliche Pole auf. Australien befindet sich in sehr zentraler Lage mit einer starken Beziehung zum Europa-zentrischen Pol. Peripherer gelegen sind die anderen zwei Pole. Die Lage Südafrikas bestätigt seine besondere doppelte Rolle zwischen südlichem Afrika und Ozeanien.

Die OST-/SÜDOSTASIATISCHE Region ist die einzige, die sich auf zwei Hauptpole verteilt (Abbildung 36): Japan gehört zum Europa-zentrischen Hauptpol und China und Südostasien mit der südasiatischen Region zum Asien-zentrischen Hauptpol. SÜDASIEN und die PHILIPPINEN sind mit der grösseren OST-/SÜDOSTASIATISCHEN Region eng verflochten und befinden sich im topologischen Raum im selben Gebiet.

Abbildung 36: Regionen des globalen Städtenetzes im topologischen Raum: Ausschnitt Asien



Diese drei Regionen haben auch einen sehr hohen Anteil an intraregionalen Beziehungen zwischen den verschiedenen Polen. Der Grund dafür ist, dass zentrale Städte sich auf den beiden Hauptpolen befinden (Südostasien / China / Indien und Westeuropa / Japan / Nordafrika), Städte, die zu deren *hinterworld* gehören, sich hingegen auf peripheren und kleinen Polen (Kerala, Westbengala, Korea, usw.) befinden. Die Beziehungen zwischen den Hauptpolen und den peripheren Polen sind auch mehrheitlich sehr

schwach. Ausnahme ist Korea: alle koreanische Städte inklusive Seoul befinden sich in einem gemeinsamen Pol mit einer starken Hauptbeziehung zum Europa-zentrischen Hauptpol. Die schwache Kohäsion der drei Regionen wird mit einem tiefen Modularitätswert (0.06) bestätigt.

Aus der Betrachtung der einzelnen Regionen lassen sich folgende Schlüsse ziehen:

- Europa, Nordamerika, Australien, Teile des Mittelmeerraums und die Mehrheit von Asien bilden auf regionaler Ebene ein zentrales Gebiet.
- Der Süden der Welt und einzelne Städte Asiens sind hingegen peripher. Brasilien und Russland übernehmen eine Zwischenstellung.
- Während die meisten Regionen, insbesondere die angelsächsischen Regionen, sich auf ein Gebiet im topologischen Raum konzentrieren, fällt AFRIKA/NAHER OSTEN wegen der schwachen interregionalen Beziehungen, der tiefen Modularität und der hohen Fragmentierung auf. Wenn man Afrika als geographischen Kontinent betrachtet, ist die Fragmentierung noch ausgeprägter.

Städtische Perspektive

Wenn man sich nun von den Regionen abwendet, rücken die Zentralitäten der einzelnen Städte ins Zentrum. Der gewichtete *degree*¹⁰ ist ein lokales Mass und gibt die Zentralität in Bezug auf die Nachbarstädte wieder. Die nähebasierte und die *betweenness* Zentralität sind globale Masse; das erste Mass betrachtet die Erreichbarkeit, das zweite die Kontrollmöglichkeit eines Knotens im Netzwerk (siehe Kapitel 5.2). Alle drei Masse zeigen eine ausgesprochene sogenannte *long-tail*-Verteilung. Allerdings weisen bei Betrachtung der Städte mit Zentralitätswerten über dem Median die *betweenness* Zentralität die stärkste und die nähebasierte Zentralität die schwächste hierarchische Struktur auf (gemessen an der Steigung der Log-Log-Kurve). Das gleiche hierarchische Muster, stark ausgeprägt für die *betweenness* Zentralität und schwach für die nähebasierte Zentralität, zeigt sich auch bei der Untersuchung der Spitze der Hierarchie (Tabelle 23). Der Zentralitätswert für Los Angeles, das bei der *betweenness* Zentralität auf Rang zehn liegt, beträgt 26.87 (von 100) und 94.14 für Wien, das bei der nähebasierten Zentralität auf Rang zehn liegt.

Die Spitze dieses globalen Städtennetzes tritt ausgehend vom Zentralitätswert klar hervor: London, New York und Paris rangieren, abgesehen von der *betweenness* Zentralität, mit deutlichem Abstand in den Werten auf den ersten drei Rängen. Bei der *betweenness* Zentralität wird das städtische Triumvirat um Peking (Rang 2) erweitert. London ist mit dem ersten Rang bei allen Zentralitäten die absolute Spitze des Städtesystems.

¹⁰ Wegen der sprachlichen Verzerrung wird der gewichtete *degree* nicht nur für das ganze Netzwerk, sondern auch für den MST berechnet.

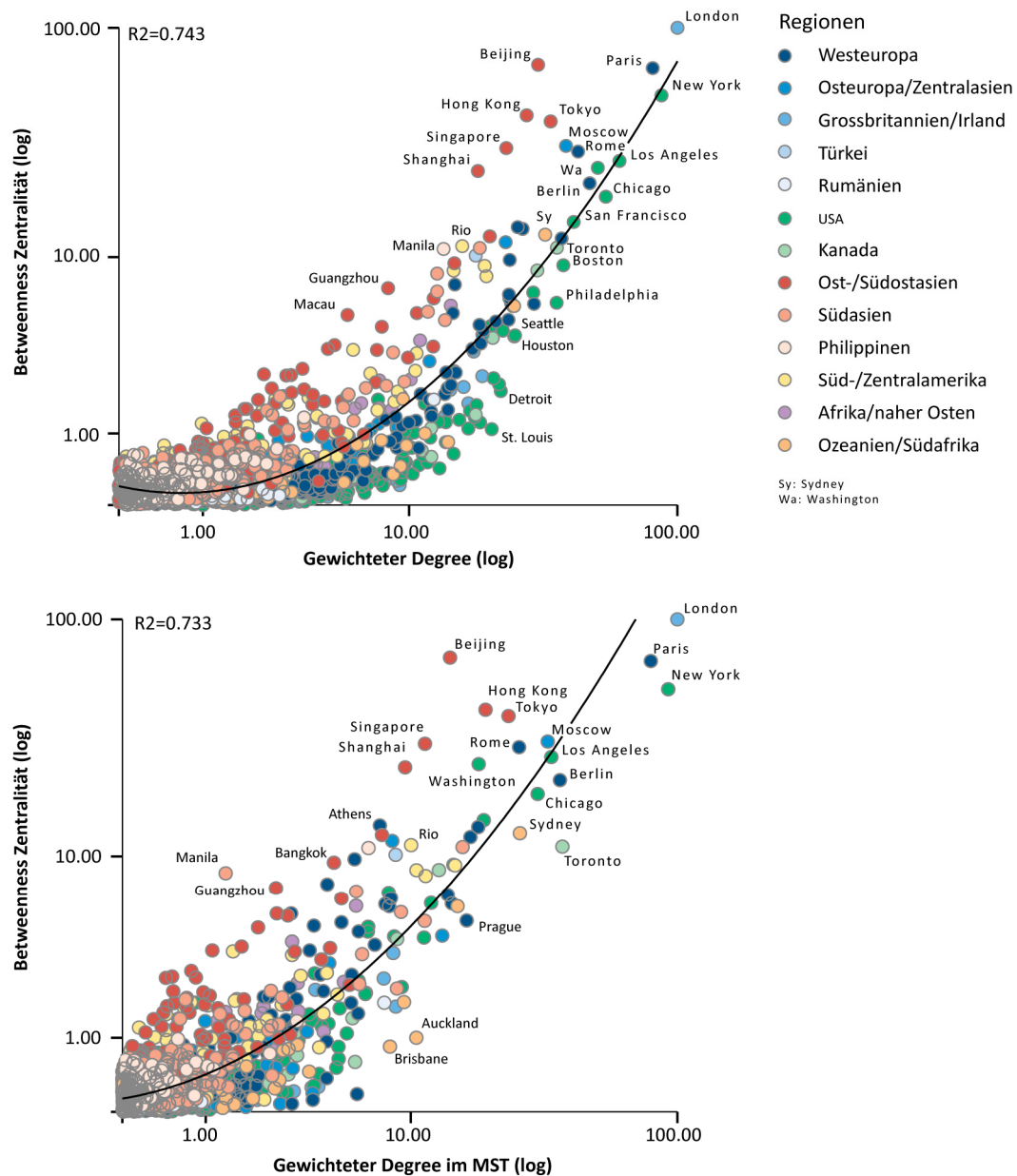
Tabelle 23: Zentralitäten (normalisierte Werte mit Rangwerten in Klammern) der zentralsten Städte

Stadt	Region	Gewichteter <i>degree</i>	Gewichteter <i>degree</i> (MST)	Nähebas. Zent.	<i>Betweenness</i> Zentralität
London	Grossbritannien/ Irland	100.00(1)	100.00(1)	100.00(1)	100.00(1)
New York	USA	87.47(2)	92.64(2)	99.08(3)	51.56(4)
Paris	Westeuropa	81.31(3)	79.96(3)	99.52(2)	67.39(3)
Toronto	Kanada	36.32(13)	37.84(4)	89.34(19)	11.05(23)
Berlin	Westeuropa	47.90(7)	37.03(5)	96.43(7)	21.46(13)
Los Angeles	USA	61.58(4)	34.40(6)	97.89(4)	26.87(10)
Moscow	Osteuropa/ Zentralasien	39.21(10)	33.37(7)	93.74(11)	31.22(7)
Chicago	USA	54.86(5)	30.62(8)	97.63(5)	18.72(14)
Sydney	Ozeanien/ Südafrika	32.86(16)	26.28(9)	89.27(20)	12.67(18)
Rome	Westeuropa	43.44(8)	26.08(10)	95.59(8)	29.54(9)
Tokyo	Ost-/Südostasien	34.44(15)	23.80(11)	90.29(15)	39.87(6)
Hong Kong	Ost-/Südostasien	28.03(21)	19.48(12)	89.56(17)	42.35(5)
San Francisco	USA	41.83(9)	19.18(13)	94.61(9)	14.45(15)
Washington	USA	51.28(6)	18.35(14)	96.65(6)	25.08(11)
Vienna	Westeuropa	37.78(12)	17.06(15)	94.14(10)	12.18(20)
Beijing	Ost-/Südostasien	30.94(17)	14.24(16)	90.61(14)	69.65(2)
Singapore	Ost-/Südostasien	23.55(33)	11.39(17)	87.42(26)	30.55(8)

Der Vergleich der Zentralitätswerte, insbesondere des gewichteten *degrees* und der *betweenness* Zentralität (Abbildung 37) hebt besondere Wertanomalien hervor, die ein erstes Indiz für bestimmte Rollen einzelner Städte darstellen. Das obere Diagramm stellt der *betweenness* Zentralität den gewichteten *degree* im Gesamtnetzwerk gegenüber, während das untere Diagramm den gewichteten *degree* im MST darstellt. Städte mit einer ausgeprägten Zentralitätsanomalie platzieren sich in Abbildung 37 mit einer grossen Abweichung von der Regressionsgerade. Insbesondere asiatische Städte (Peking, Hong Kong, Singapur, Shanghai, Tokyo, Guangzhou, Macau) weisen eine abweichend hohe *betweenness* Zentralität auf. Sie übertrifft die Erwartungen, welche man aufgrund des gewichteten *degrees* haben kann. Für Paris, Moskau und Rom sind die Zentralitätsanomalien weniger ausgeprägt. Städte mit über den Erwartungen liegender *betweenness* Zentralität nehmen eine Brückenfunktion im Netzwerk ein und verbinden unterschiedliche Regionen miteinander. Für Chicago, Boston, Philadelphia, Detroit und St. Louis ist die *betweenness* Zentralität kleiner als erwartet. Der Vergleich zwischen den zwei Diagrammen zeigt, dass die *betweenness* Zentralität stärker vom gewichteten *degree* im Gesamtnetzwerk als vom MST abweicht. Die Anomalien werden durch die Filterung des Netzwerkes teilweise ausgeglichen. Trotzdem sind die Muster in den beiden Diagrammen ähnlich: OST-

/SÜDOSTASIATISCHE und PHILIPPINISCHE Städte haben eine positive Anomalie, die WESTEUROPÄISCHEN und insbesondere die US-AMERIKANISCHEN und KANADISCHEN Städte haben negative Anomalien.

Abbildung 37: Zentralitätsanomalien zwischen *betweenness* Zentralität und gewichtetem *degree* der Städte nach Guimerà et al. (2005)



Aus der Betrachtung der Zentralitäten lässt sich zudem schliessen, dass etwa dreissig Städte aus USA, KANADA, WESTEUROPA, OST-/SÜDOSTASIEN, sowie OSTEUROPA/ZENTRALASIEN und OZEANIEN/SÜDAFRIKA herausragende Werte aufweisen. Die asiatischen Städte heben sich insbesondere durch ihren globalen Einfluss (*betweenness* Zentralität) und ihre Brückenfunktion und die nordamerikanischen Städte durch ihren lokalen Einfluss (gewichteter *degree*) hervor. Die kontinentaleuropäischen Städte weisen ein ausgeglichenes lokal-global-Profil aus. London ragt über alle Städte hinaus.

Eine reine Betrachtung der Zentralitäten wäre für eine vertiefte Identifikation der globalen Städte, als Spitze und Kern des globalen Städtensetzes, unangemessen. Die universellen Rollen nach Guimerà und Amaral (2005a) ermöglichen eine schlüssigere Einschätzung und eine differenziertere Trennung zwischen den globalen Städten und den restlichen, bevölkerungsmässig grossen, national und regional relevanten Städten, die jedoch ohne entscheidende globale Stellung im Städtensetz sind. Die Rollendefinition schliesst neben der lokalen Zentralität, die regionale Organisation, die Heterogenität und Stärke der globalen Einbettung mit ein. Konkret werden die Rollen (Abbildung 38) durch den gewichteten intraregionalen *degree* (Z-Wert) und den Partizipationskoeffizienten (P-Wert) definiert. Dank dem Z-Wert werden Hubs von Nicht-Hubs unterschieden. Hubs kommen globalen Städten, also dem Kern des Städtensetzes gleich. Von den 3'532 Städten sind nur 95 Hubs, dies entspricht einem Anteil von 2.7%. Davon sind 37 (38.9%) verwandtschaftslose (R7) und 58 (61.1%) verbindende Hubs (R6). Die meisten Städte (2'537) im Städtensetz sind peripher. Die kleinen Anteile an verwandtschaftslosen Knoten und verwandtschaftslosen Hubs bestätigen die modulare Struktur des Netzwerkes (Guimerà et al. 2005).

Abbildung 38: Verteilung der Rollen im Städtensetz

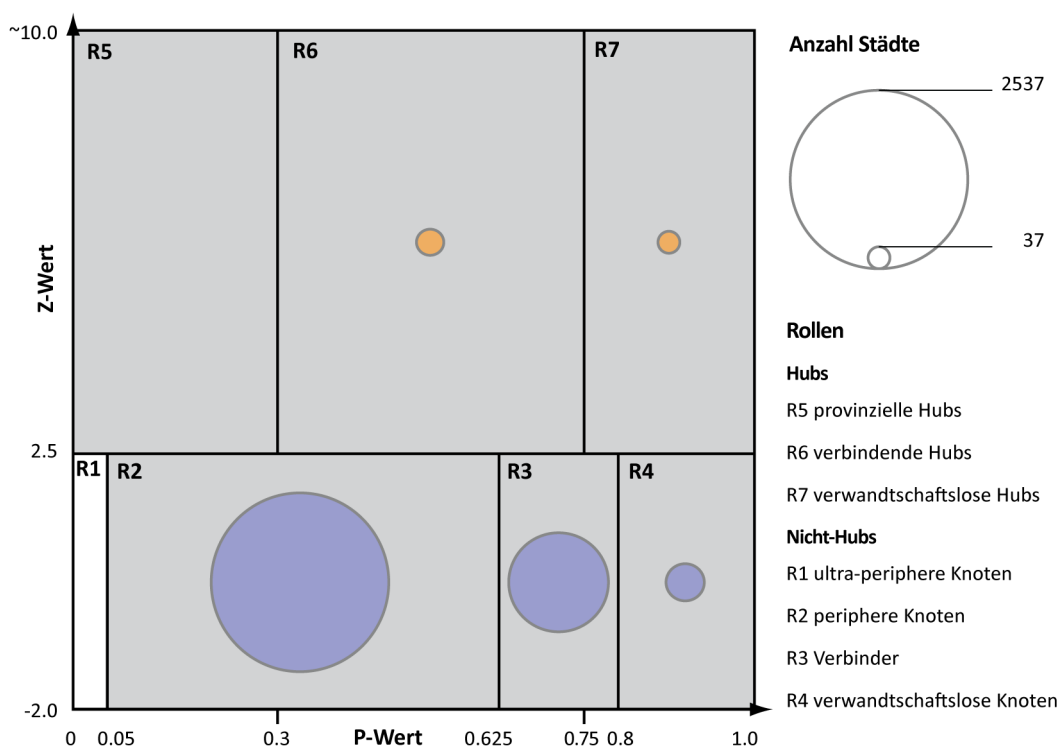


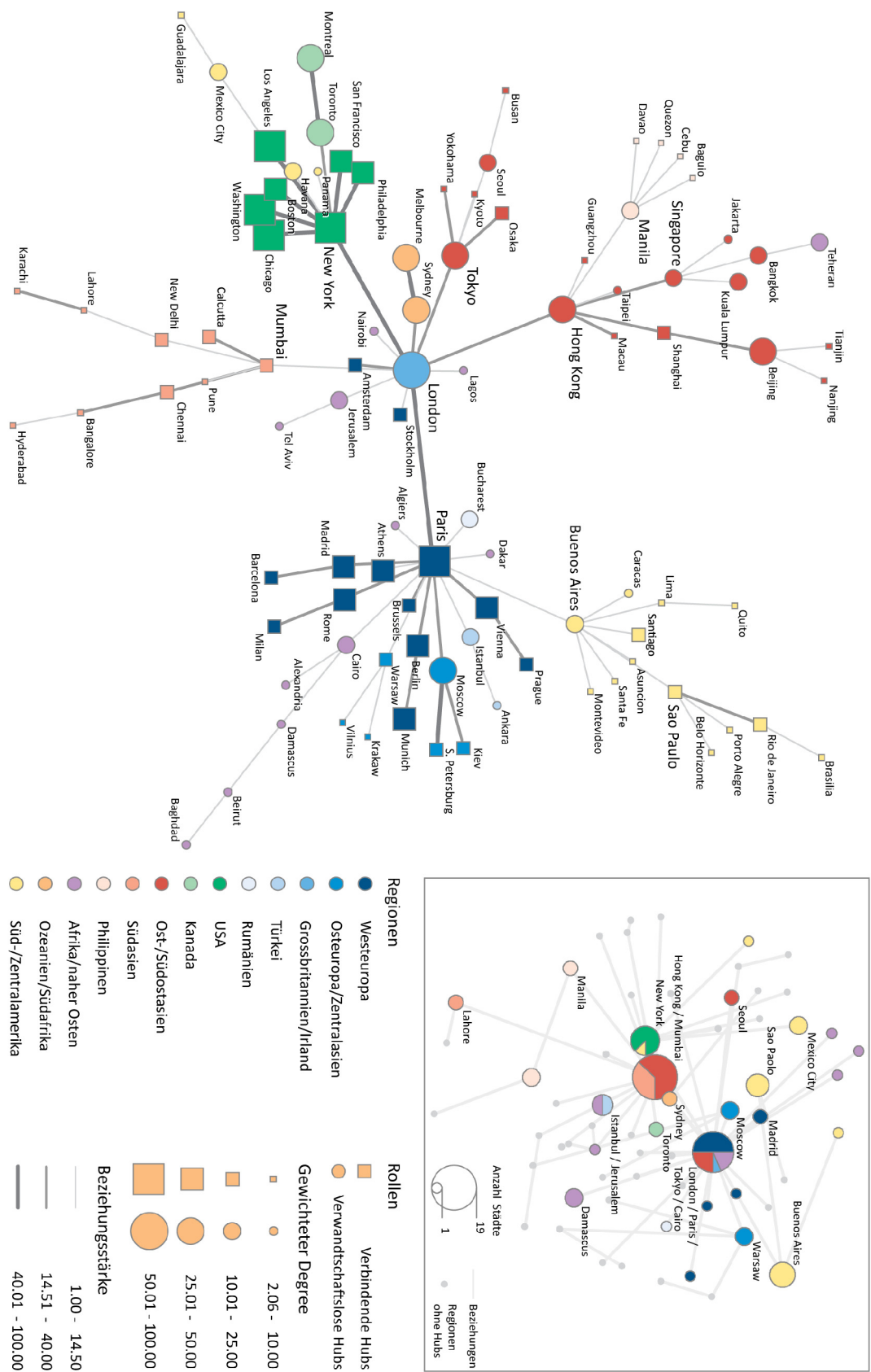
Abbildung 39¹¹ stellt die globalen Städte und ihre wichtigsten Beziehungen in der Form des MST dar. Dank der Übersichtskarte lassen sich die globalen Städte in der regionalen Ansicht wiederfinden. Daraus wird ersichtlich, dass die globalen Städte auf weniger als

¹¹ Anhang 2 entspricht Abbildung 39, jedoch wird nicht der gewichtete *degree* sondern die *betweenness* Zentralität hervorgehoben.

die Hälfte (27 von 63) der Pole konzentriert sind. 24 der globalen Städte gehören zu den drei Hauptpolen.

Das System der globalen Städte ist mehrpolig, wobei wiederum drei Hauptzentren erkennbar sind: Paris, London und New York. London und Paris befinden sich auf dem Europa-zentrischen und New York auf dem US-zentrischen Hauptpol. Zwischen diesen drei Städten bestehen die stärksten interregionalen Beziehungen. Die Interaktion Paris-New York ist nicht dargestellt, ist in der Intensität jedoch vergleichbar mit der Beziehung London-New York.

Abbildung 39: Globales Städtenetz zwischen den Hubs und deren Lage in der regionalen Ansicht



Paris und London haben eine hohe lokale und globale Zentralität und sind zudem, bei der Betrachtung der Hauptbeziehungen, Zentren eines grossen *hinterworld*. New York weist hingegen hohe Zentralitäten aus, obwohl das *hinterworld* sich auf die nord- und zentralamerikanischen globalen Städte beschränkt. Paris kristallisiert sich als Hauptzentrum der kontinentaleuropäischen, russischen, arabischen und südamerikanischen Städte heraus. Buenos Aires und in zweiter Linie Sao Paolo sind die Zentren Südamerikas. Die europäischen Städte sind entweder direkt mit Paris verbunden oder mit der Hauptstadt des jeweiligen Landes. Ausnahmen sind Prag und Kiew, welche über Wien bzw. Moskau mit Paris verbunden sind.

Amsterdam und Stockholm sind die einzigen europäischen Städte mit Hauptbeziehungen zu London. London hebt sich insbesondere als Zentrum eines asiatisch-ozeanisch geprägten *hinterworld* hervor. Hinzu kommen die israelischen Städte und die Städte aus der Sub-Sahara. Global betrachtet kann London durchaus als Zentrum des gesamten Systems, bzw. als DIE globale Stadt betrachtet werden.

Die OST-/SÜDASIATISCHEN Städte organisieren sich wiederum in eine japanisch-koreanische und eine chinesisch-südasiatische Region. Tokyo ist das Zentrum der ersten, Hong Kong der zweiten. Subzentren der chinesisch-südasiatischen Region sind Singapur, Peking und Manila. Dies zeigt die engen Beziehungen der Städte der philippinischen zu den Städten der chinesisch-südasiatischen Region.

Die regionale Organisation spiegelt sich in den Beziehungen zwischen den globalen Städten sehr gut wider. So ist Sao Paolo nur indirekt über Buenos Aires mit Paris verbunden, so wie die brasilianische Region nur indirekt über die südamerikanische Region mit dem Europa-zentrischen Hauptpol verbunden ist. Ein Pol hat sich jedoch bei der Betrachtung der globalen Städte aus regionaler Perspektive aufgeteilt. Die türkischen und israelischen Städte sind auf regionaler Ebene Teil eines gemeinsamen Pols. Jerusalem und Tel Aviv beziehen sich nun auf London, Istanbul und Ankara auf Paris, engere gegenseitige Verbindungen sind nicht erkennbar. Der Grund für das Zusammenrücken auf regionaler Ebene könnte somit eher ein ähnliches Beziehungsmuster zu den restlichen Städten sein, insbesondere in Asien, USA und Europa, als enge gegenseitige Beziehungen.

Auf der Ebene der Beziehungen sind neben der Hauptachse Paris-London-New-York vor allem intraregionale Beziehungen in WESTEUROPA, USA, KANADA, und OZEANIEN/SÜDAFRIKA sehr intensiv. In der OST-/SÜDOSTASIATISCHEN Region ist die Lage heterogener: während Hong Kong, Tokyo, Peking, Singapur und Schanghai durchaus auch starke Beziehungen haben, sind die restlichen Beziehungen schwach. Abbildung 40 zeigt das regionale Spektrum dieser Beziehungen auf.

Verbindende Hubs (R6) weisen einen hohen Anteil, meistens um oder über 50%, an intraregionalen Beziehungen auf. Ihre Zentralität ist insbesondere durch die starke regionale Einbettung gegeben. Sie verbinden die periphereren Städte ihrer Region mit dem Rest des Netzwerkes. Verbindende Hubs konzentrieren sich in den grossen Regionen

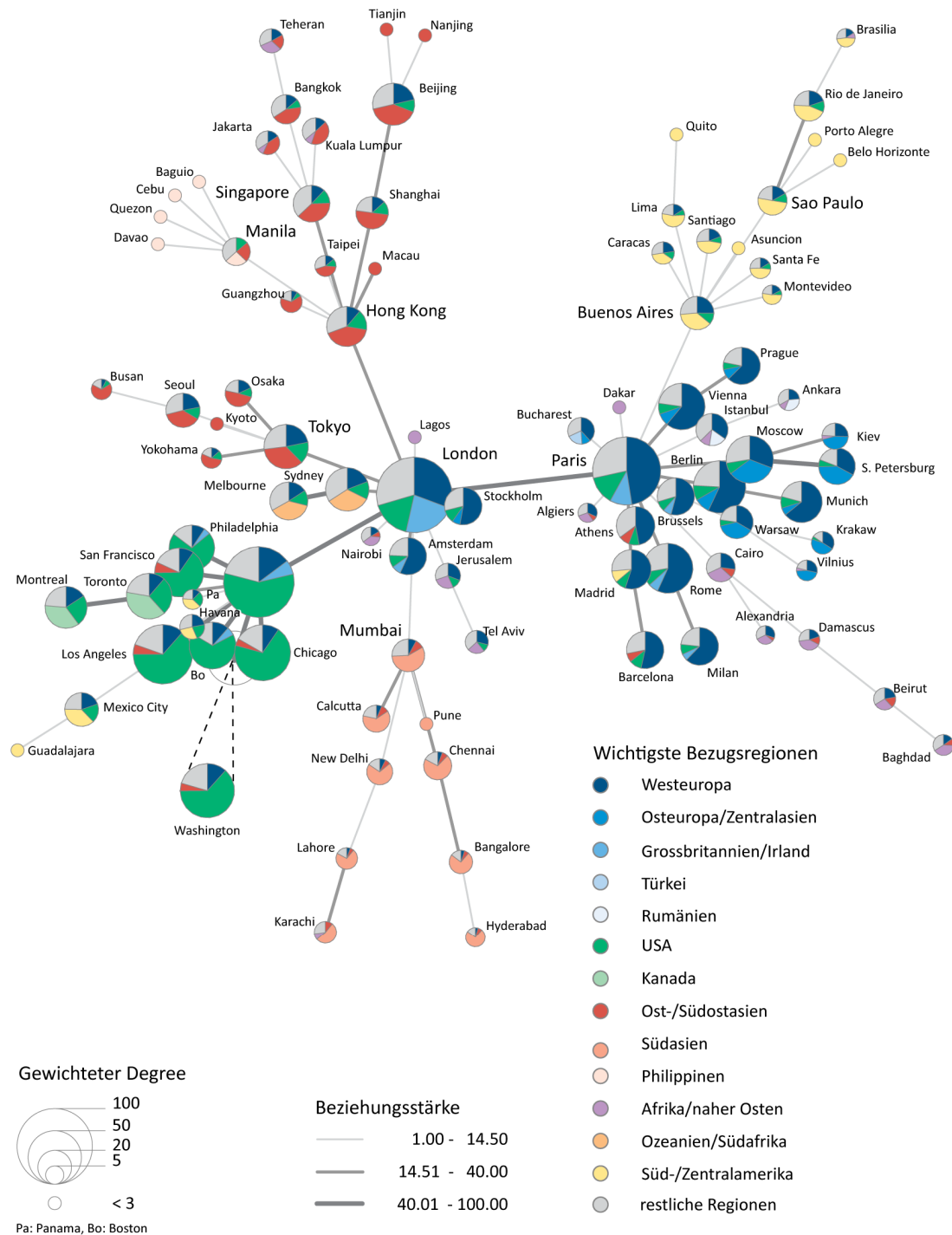
mit hoher Modularität und allgemein hoher Zentralität der Städte: WESTEUROPA und die USA.

Verwandschaftslose Hubs (R7) haben einen geringeren regionalen Rückhalt. Dies ist insbesondere für kleine Regionen der Fall (OZEANIEN/SÜDAFRIKA, GROSSBRITANIEN/IRLAND, KANADA, RUMÄNIEN und TÜRKEI). Globale Städte in grösseren Regionen mit einer geringen Kohäsion (tiefe Modularitätswerte) sind mehrheitlich auch verwandschaftslose Hubs (R7). Dies zeigt sich sehr gut bei der Region AFRIKA/NAHER OSTEN. Dieses Muster trifft nicht auf die südasiatische Region zu: die globalen Städte sind ausnahmslos verbindende Hubs.

In der OST-/SÜDOSTASIATISCHEN und der OSTEuropÄISCH/ZENTRALASIATISCHEN Region ist die Lage geteilt: während die zentralen Städte, insbesondere in Bezug auf die *betweenness* Zentralität, verwandschaftslos und somit interregional vernetzt sind, zählen die periphereren Städte zu den verbindenden Hubs.

Abbildung 40 verdeutlicht nicht nur den unterschiedlichen Anteil an intraregionalen Beziehungen und somit den Unterschied zwischen den Hubs, sondern zeigt das intra- und interregionale Beziehungsmuster jeder Stadt. Das Muster unterscheidet sich insbesondere zwischen den Regionen, zeigt aber auch kleinere Differenzen innerhalb der Regionen.

Abbildung 40: Beziehungen der globalen Städte nach Regionen



Trotz erheblicher regionaler Unterschiede pflegen die meisten Städte starke Beziehungen zur WESTEUROPÄISCHEN Region und den USA. Ausnahmen sind die AFRIKANISCH/NAHÖSTLICHEN, sowie die SÜDASIATISCHEN Städte, welche insbesondere mit WESTEUROPÄISCHEN Städten, und die PHILIPPINISCHEN Städte, welche insbesondere mit den USA verbunden sind.

Die WESTEUROPÄISCHEN Städte teilen sich aufgrund der wichtigsten regionalen Beziehungen in vier Untergruppen auf. Alle haben wichtige interregionale Beziehungen zu den USA, unterscheiden sich jedoch in der weiteren Beziehung:

- Die globalen Städte Ost- und Nordeuropas beziehen sich auch stark auf die benachbarte OSTEUPÄISCH/ZENTRALASIATISCHE Region,
- Die westlichen Städte orientieren sich hingegen stärker nach GROSSBRITANNIEN/IRLAND,
- Madrid bezieht sich stark auf SÜD-/ZENTRALAMERIKA und
- Barcelona und Athen beziehen sich auf OST-/SÜDOSTASIEN.

Vergleichbar ist das Muster für die US-AMERIKANISCHEN Städte: während New York, Boston und Philadelphia sich stärker auf GROSSBRITANNIEN/IRLAND beziehen, orientieren sich die restlichen Städte mehr Richtung OST-/SÜDOSTASIEN.

Obwohl sich die OST-/SÜDOSTASIATISCHEN und die ZENTRAL-/SÜDAMERIKANISCHEN Städte in jeweils zwei Untergruppen teilen, weisen sie sehr homogene Muster auf. Beide Regionen beziehen sich auf interregionaler Ebene auf WESTEUROPA und USA. Grosse Unterschiede innerhalb der Region sind wiederum in der AFRIKANISCH/NAHÖSTLICHEN Region festzustellen: die israelischen globalen Städte differenzieren sich klar von den restlichen. Jerusalem und Tel Aviv beziehen sich vor allem auf WESTEUROPA und die USA, die restlichen Städte hingegen auf OST-/SÜDOSTASIEN und WESTEUROPA.

Aus städtischer Perspektive lassen sich folgende Schlüsse ziehen:

- Die Städte unterscheiden sich stark in ihrer lokalen und globalen Zentralität. Dabei sind London, Paris und New York besonders zentral.
- Die Rollendefinition bringt 95 globale Städte hervor. Paris, London und New York übernehmen wiederum eine herausragende Position. Insbesondere London kann als die globalste Stadt betrachtet werden.
- Die intraregionale Fragmentierung, insbesondere der AFRIKANISCH/NAHOSTASIATISCHEN Region, aber auch der OST-/SÜDOSTASIATISCHEN und SÜD-/ZENTRALAMERIKANISCHEN Region, ist auch zwischen den globalen Städten festzustellen.
- WESTEUROPA und die USA sind die wichtigsten interregionalen Bezugsregionen für die meisten Städte.

7.2 Funktionen im Städtenetz

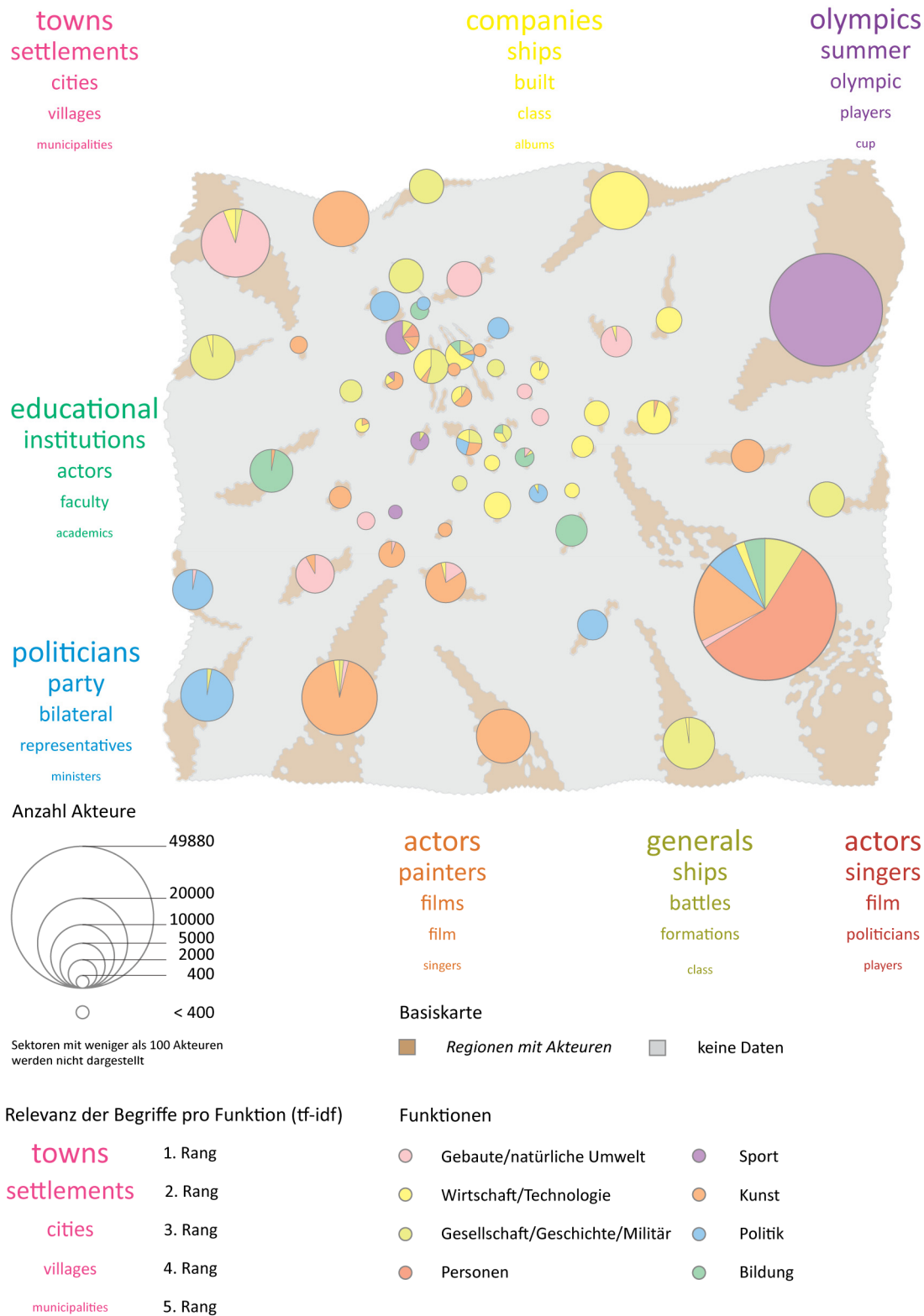
Die Beziehungen zwischen Städten sind das Resultat des Handelns verschiedener Akteure, seien es Personen oder Institutionen. Diese setzen Flüsse an Personen, Informationen und Güter zwischen den Städten in Bewegung (Smith und Timberlake 1995b). Der Kontext, in dem die Interaktionen und deren Funktion entstehen, ist sehr unterschiedlich. Smith und Timberlake (1995b) identifizieren auf theoretischer Ebene vier Funktionen: ökonomische, politische, kulturelle und soziale Funktionen. Dieses Kapitel zeigt die Funktionen, zu denen die Akteure im untersuchten Städtenetz gruppiert werden können. In Kapitel 6.3 werden die Funktionen auf ihre Robustheit evaluiert. Zudem zeigt das Kapitel, dass die Funktionen¹² grob mit den Begriffen GEBaute/NATÜRLICHE UMWELT, WIRTSCHAFT/TECHNOLOGIE, SPORT, PERSONEN, GESELLSCHAFT/-GESCHICHTE/MILITÄR, KUNST, POLITIK und BILDUNG umschrieben werden können.

Abbildung 41 ist eine Überarbeitung der Abbildung 31 und zeigt die Verteilung der acht identifizierten Funktionen in einem mit SOM generierten topologischen Raum. Funktionen werden mit Farbtönen hervorgehoben. Aus der Darstellung wird auch ersichtlich, dass der Raum in 57 abgetrennte Regionen (braun) unterteilt ist. Auf diesen Regionen platzieren sich die Akteure. Dazwischen verteilen sich grosse Teile des topologischen Raumes, in denen keine Akteure vorkommen (grau). Die Akteure einer Region gehören mehrheitlich einer Funktion (Farbe) an. Die Kreisdiagramme geben nun die Anzahl Akteure pro Region und die Zuteilung zu den Funktionen wieder. Ein Lesebeispiel: die Region oben rechts umfasst viele Akteure (Grösse), die ausschliesslich der Funktion SPORT zugehören. Der topologische Raum zeigt zudem ein ausgeprägtes Muster: grosse, gemessen an der Anzahl Akteure, klar voneinander abgegrenzte Regionen bilden den Rand des topologischen Raumes und kleine, teilweise sehr nahe beieinander liegende Regionen bilden das Zentrum des Raumes. Zudem sind einige der kleinen Regionen hinsichtlich der funktionalen Zusammensetzung sehr heterogen (mehrere Sektoren in einem Diagramm). Grundsätzlich stellen die grossen Regionen am Rande die thematischen Schwerpunkte im Datensatz dar. Die kleinen Regionen in der Mitte sind hingegen die thematischen Restmengen. Wegen der Struktur des SOM weisen die Regionen in der Mitte tendenziell eine grössere Ähnlichkeit zu mehreren anderen Regionen auf. Inhaltlich handelt es sich dabei um mehrdeutige Inhalte. Aus der Untersuchung der Ähnlichkeit zwischen den Regionen hat sich jedoch gezeigt, dass nicht alle kleinen Regionen diesem Muster entsprechen. Wenige Regionen in der Mitte des Raumes sind besonders klein und haben sehr geringe Ähnlichkeiten mit allen anderen Regionen. Es könnte sich dabei um thematische Nischen handeln. Wegen der geringen Grösse und Ähnlichkeit sind diese Akteure schwierig im topologischen Raum zu platzieren. Aus der Kombination der Klassierung mit dem Blondel-Algorithmus (Farben) und der Geometrie des topologischen Raumes lässt sich in dieser Hinsicht schliessen, dass in der Tendenz die grossen Regionen am Rande des topologischen Raumes den inhaltlichen Schwerpunkt

¹² Funktionen und Subfunktionen werden im Text mit KAPITÄLCHEN geschrieben.

der jeweiligen, mit Farbe hervorgehobenen, Funktionen sind. Die kleinen Regionen in der Mitte werden hingegen der Funktion, zu der sie die grösste Ähnlichkeit aufweisen, zugerechnet. Sie bilden dabei kleinere, dem Funktionsschwerpunkt inhaltlich teilweise ähnliche Themengebiete.

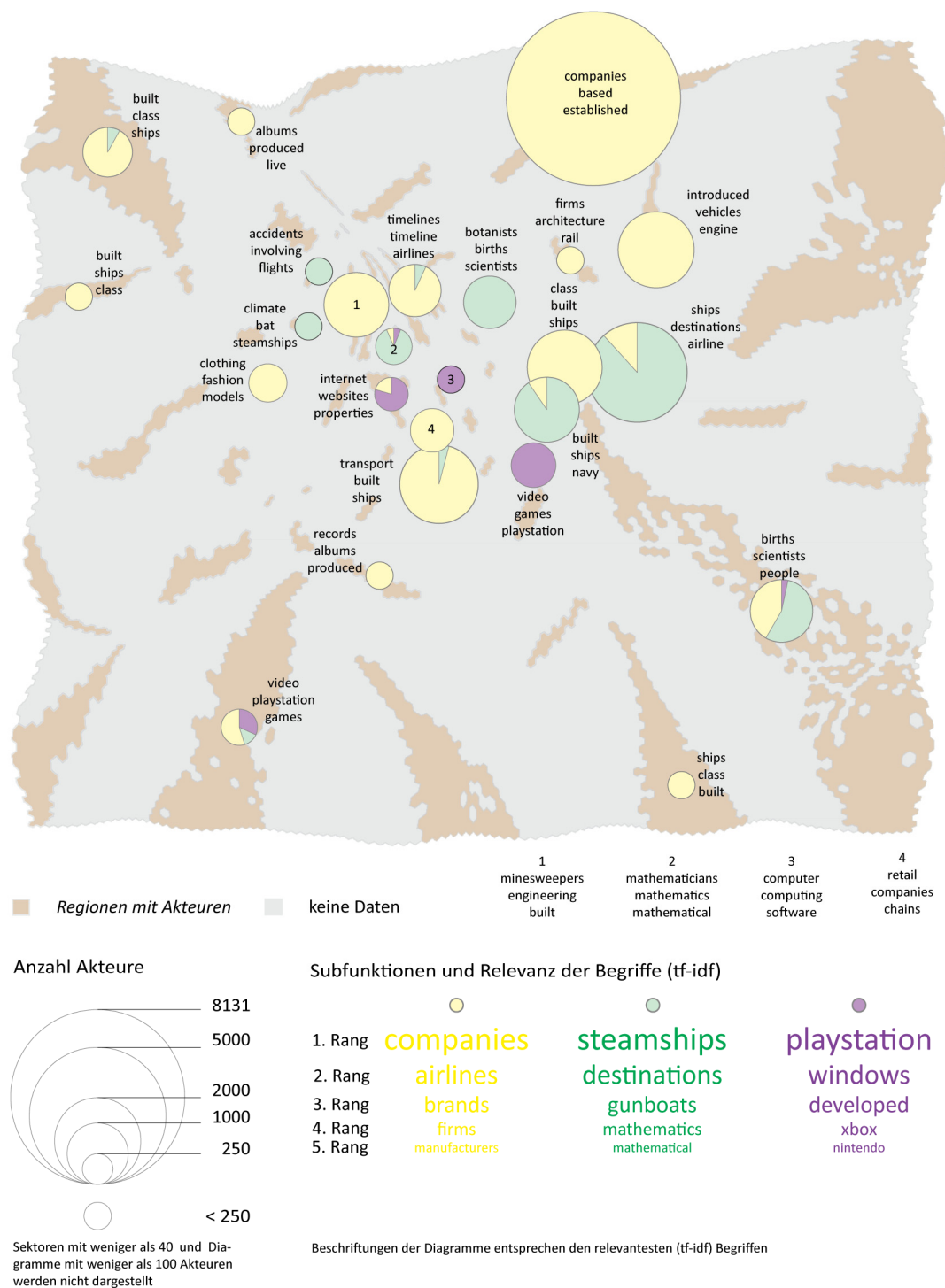
Abbildung 41: Funktionen im Städtenetz und deren Grösse



Wenn man Abbildung 41 aus der Perspektive der acht Funktionen betrachtet, sieht man, dass sie sich nicht nur hinsichtlich des Inhalts (Wortwolken), sondern auch der Grösse (Grösse der Kreise) und der inneren Heterogenität (Anzahl und Distanz zwischen den Regionen) differenzieren. Die grösste Funktion ist mit 49'883 Akteuren KUNST, die kleinste BILDUNG (11'321 Akteure). Die weiteren variieren zwischen 21'000 (POLITIK) und 34'000 (SPORT). Da der topologische Raum der Distanz-Ähnlichkeitsmetapher folgt, ist die räumliche Konzentration ein Hinweis auf die Homogenität der Funktionen. In dieser Hinsicht ist die Funktion PERSONEN sehr homogen: fast 100% der Akteure konzentrieren sich auf die Region unten rechts. Eine gewisse Heterogenität aufgrund der grossen räumlichen Ausdehnung ist trotzdem zu erwarten. Für die Funktion WIRTSCHAFT/TECHNOLOGIE lassen die Verteilung über 25 Regionen im topologischen Raum, sowie die hohe Anzahl an kleinen Regionen in der Mitte des Raumes auf eine hohe inhaltliche Heterogenität und auf mehrere, kleinere inhaltliche Schwerpunkte schliessen. Diese graphische und inhaltliche Multipolarität drängt eine vertiefte Analyse jeder einzelnen Funktion auf, die nun in den nächsten Abschnitten gemacht wird.

Abbildung 42 zeigt mit der gleichen Darstellungsart wie in Abbildung 41 die Akteure, welche zur Funktion WIRTSCHAFT/TECHNOLOGIE gehören. Die Farbtöne stellen die drei Subfunktionen, welche auf einem tieferen Aggregationsniveau bzw. bei höherem Detaillierungsgrad erkennbar sind, dar. Die Heterogenität zeigt sich de facto in unterschiedlichen Aspekten. Die Beschriftungen entsprechen den drei relevantesten (tf-idf) Begriffen für die Akteure der Funktion WIRTSCHAFT/TECHNOLOGIE der jeweiligen Region.

Abbildung 42: Funktionen im Städtenetz und deren Grösse: Auszug WIRTSCHAFT/TECHNOLOGIE



Diese Informationen werden durch Abbildung 43 ergänzt. Diese zeigt mit unterschiedlichen Farben die latenten Themen, aus welchen sich die Funktionen ergeben. Die Verteilung im topologischen Raum der latenten Themen ist jedoch nicht diskret, sondern kontinuierlich. Die hervorgehobenen Bereiche entsprechen der maximalen Ausprägung. Grau schraffiert sind die 57 Regionen aus Abbildung 42 eingetragen, auf denen sich die Akteure verteilen. Durch das Überlagern der zwei Ebenen wird die thematische Verteilung der Regionen ersichtlich. Durch die Kombination der Informationen aus den bei-

den Darstellungen lässt sich der Inhalt der Funktion eruieren. Ein Lesebeispiel: Aus Abbildung 42 wird ersichtlich, dass die Region oben rechts ca. 8'100 Akteure (Grösse des Kreises) umfasst und somit die Grösste ist. Diese Akteure gehören alle zur selben Subfunktion (Farbton=gelb). Inhaltlich lassen sich diese Akteure mit den Begriffen *companies / based / established* umschreiben. Diese entsprechen den relevantesten Begriffen (tf-idf-Methode). Der unternehmerische Charakter dieser Akteure wird in Abbildung 43 bestätigt: Diese Region hat ihren Schwerpunkt im Thema *companies* (gelber Hintergrund). Das Fazit daraus ist, dass die Funktion WIRTSCHAFT/TECHNOLOGIE einen wichtigen Schwerpunkt im Bereich Unternehmertum hat.

Mit diesem Interpretationsmuster lassen sich für die Funktion WIRTSCHAFT/TECHNOLOGIE folgende Aspekte hervorheben (Abbildung 42 und Abbildung 43). Die grösste Subfunktion (gelb) lässt sich mit den Begriffen *companies / airlines / brands / firms / manufacturers* umschreiben. Diese stammen, mit Ausnahme von *airlines*, aus dem unternehmerischen Bereich der Wirtschaft. Die grosse Mehrheit der Akteure aus der erwähnten Subfunktion konzentrieren sich in der Region *companies / based / established* (oben rechts), die ihren Schwerpunkt im latenten Thema *companies* hat (Abbildung 43). Die weiteren grossen gelben Diagramme (Abbildung 42) offenbaren eine zweite thematische Facette. Relevante Begriffe sind dabei *introduced / vehicles / engine* (gelbes Kreisdiagramm oben rechts), *class / built / ships* (gelbes Kreisdiagramm in der Mitte), *transport / built / ships* (mehrheitlich gelbes Kreisdiagramm in der Mitte) und *minesweepers / engineering / built* (Kreisdiagramm 1). Diese Begriffe haben ihren Schwerpunkt bei den Themen *marketing, equipment, industry* bzw. *engineering* (Abbildung 43). Dieser industriell-ingenieurtechnische Bereich fokussiert sich in diesem Datensatz ganz klar im Bereich Transport und Fahrzeugbau. Ähnliche Begriffe sind bei mehreren kleineren Diagrammen auch wiederzufinden. Diese umfassen die transporttechnischen bzw. fahrzeugbaulichen Akteure in unterschiedlichen Kontexten. In diesem Sinn ist davon auszugehen, dass das Kreisdiagramm unten rechts (*ships / class / built*) (Abbildung 42), welches sich im Thema *military* befindet (Abbildung 43), diese Aspekte im Kontext des Militärs bzw. des Krieges umfasst.

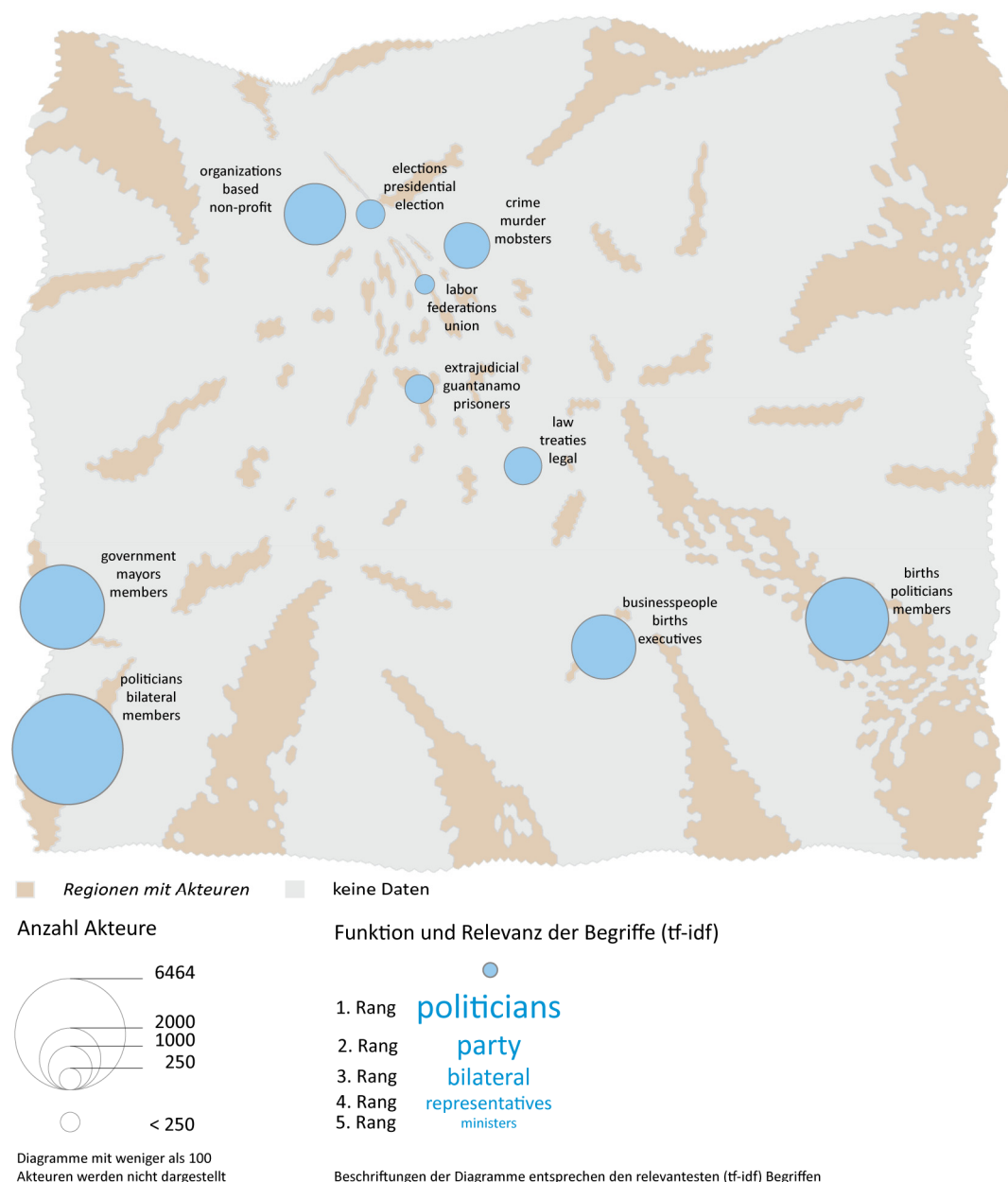
Weiter wird die Subfunktion durch den Themenbereich *commerce (retail / companies / chains)* und *clothing (clothing / fashion / models)* ergänzt.

Die zweite Subfunktion knüpft mit den Begriffen *steamships, gunboats* und, *destinations* beim Bereich Transport und Fahrzeugbau gleich an. Die Begriffe der zwei grössten Kreisdiagramme, *ships / destinations / airline* und *built / ships / navy* stimmen mit den oben erwähnten Begriffen überein. Die schwerpunktmässigen Themenbereiche sind dabei *technology* und *science*. Desweiteren kommen mit den Begriffen *mathematics* und *mathematical* der mathematische Aspekt der Wissenschaft und mit *botanists / births / scientist* die WissenschaftlerInnen dazu.

Die kleinste Subfunktion umfasst Akteure aus dem Informatikbereich. Die Unterhaltungselektronik ist mit *video / games / playstation* und *video / playstation / games* ebenso prominent vertreten.

Abbildung 44 illustriert die Funktion POLITIK. Sie ist verglichen mit der Funktion WIRTSCHAFT/TECHNOLOGIE thematisch homogener. Sie konzentriert sich dabei auf weniger Regionen und unterteilt sich nicht in Subfunktionen. Der grösste Teil der Akteure konzentriert sich auf die zwei Regionen unten links. Diese positionieren sich schwerpunktmässig in den Themenbereichen *politics* und *government*. Relevante Begriffe sind dabei *politicians* / *bilateral* / *members* und *government* / *mayors* / *members*. Mit *births* / *politicians* / *members* sind die Begriffe für die Akteure dieser Funktion denen des latenten Themas *people* sehr ähnlich. Ergänzt wird dieser politische Bereich durch die Akteure der Region *elections* / *presidential* / *election*. Eine kleinere Anzahl Akteure kann auf den juristischen Themenbereich zurückgeführt werden. Dieser befindet sich auf drei Regionen mit den Begriffen *crime* / *murder* / *mobsters*, *law* / *treaties* / *legal* und *extrajudicial* / *guantanamo* / *prisoners*. Hinzu kommen die Organisationen, insbesondere non-Profit Organisationen und die Geschäftsleute.

Abbildung 44: Funktionen im Städtenetz und deren Grösse: Auszug POLITIK



Aufgrund äquivalenter Überlegungen, wie sie für die Funktionen POLITIK und WIRTSCHAFT/TECHNOLOGIE in diesem Kapitel dargelegt werden, lassen sich auch die anderen sechs Funktionen untersuchen:

- GESELLSCHAFT/GESCHICHTE/MILITÄR (Anhang 3): Unterteilt sich in zwei Subfunktionen: MILITÄR/KONFLIKTE einerseits, sowie GESCHICHTE/GESELLSCHAFT andererseits. Teil der Subfunktion MILITÄR/KONFLIKTE ist auch die fiktionale Umsetzung davon (*debuts / series / television / terrorist*). Geschichte als Wissenschaft, als Erkundung der Vergangenheit (*historians / holocaust / historiography* und *archeological / archeology / ancient*) und als Vergangenheit in zeitlicher Hinsicht (*history / war / ships*) ist Teil der zweiten Subfunktion. Der Krieg

ist auch in diesem Themenbereich von Relevanz. Die gesellschaftlichen Aspekte fokussieren sich auf Religion, Philosophie und Tradition.

- PERSONEN (Anhang 4): Konzentriert sich fast ausschliesslich auf eine Region und umfasst Personen, die nicht anderen Funktionen zugeteilt sind.
- SPORT (Anhang 5): Der Hauptteil konzentriert sich wiederum auf eine Region und umfasst den Sport in seinen unterschiedlichen Facetten. Dabei spielt die Olympiade eine relevante Rolle. Ergänzt wird die Funktion durch Akteure aus den Themen Klubs und Events.
- GEBaute/NATÜRLICHE UMWELT (Anhang 6): Das Schwergewicht liegt bei den urbanen Siedlungsformen (*towns / cities*). Hinzu kommt die gebaute (*buildings / structures / skyscrapers* und *architects / architecture / buildings*) und die natürliche Umwelt (*environmental / environment / climate* und *hurricanes / hurricane / season*), sowie die Landwirtschaft. Unerwartet ist, dass Akteure aus dem Bereich Musik mit einbezogen werden.
- KUNST (Anhang 7): Musik (*group / musical / music*), Literatur (*writers / poets*), bildende (*painters*) und darstellende Kunst (*films / actors*), sowie die Unterhaltungsmedien (*radio, television*) sind in dieser Funktion zusammengefasst.
- BILDUNG (Anhang 8): Unterteilt sich in zwei Themenbereiche: einerseits die Bildung, mit dem Schwerpunkt auf der universitären Bildung (*university / alumni / faculty, alumni / university / births*) und andererseits die Gesundheit. Die Gesundheit umfasst neben der Ernährung (*cuisine / health / dishes*), die Psychologie (*psychology / psychologists / cigar*) und den klinischen Bereich (*physicians / medical / hospitals*).

Neben der absoluten Anzahl Akteure zeigt auch der Vergleich dieser Werte mit einem Datensatz, bestehend aus zufällig ausgewählten Artikeln, die Wichtigkeit der Funktionen. Daraus lässt sich abschätzen, welche Funktionen eine überproportionale Relevanz im Städtenetz haben. Tabelle 24 legt diesen Vergleich dar, wobei für jede Funktion und Subfunktion die Abweichung der relativen Häufigkeit der Akteure bezogen auf den Zufallsdatensatz berechnet wurde. Dies zeigt, dass die Funktion PERSONEN nicht nur aufgrund der absoluten Häufigkeiten von grosser Relevanz ist, sondern auch und insbesondere wegen der überproportionalen, doppelt so hohen Häufigkeiten im Bezug auf den Zufallsdatensatz. Überproportional ist auch die Häufigkeit der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR, insbesondere der Subfunktion MILITÄR/KONFLIKTE (+57.31 %). Die Funktion WIRTSCHAFT/TECHNOLOGIE zeigt ein dichotomes Bild. Während die Subfunktionen COMPUTING (-70.50 %) und TECHNOLOGIE/WISSENSCHAFT (-49.67 %) eine tiefere Bedeutung haben, ist die Subfunktion WIRTSCHAFT (+34.46 %) überproportional häufig vertreten. Die Funktion GEBaute/NATÜRLICHE UMWELT ist am stärksten untervertreten. POLITIK, SPORT, KUNST und BILDUNG zeigen nur geringfügige Unterschiede.

Tabelle 24: Vergleich der absoluten und relativen Häufigkeiten von Akteuren pro Funktion und Subfunktion zwischen einem Datensatz mit zufällig ausgewählten Artikeln und dem Städtenetz

Funktion	Subfunktion	Zufallsdatensatz		Städtenetz		Abweichung (%)
		Häufigkeit	Relative Häufigkeit	Häufigkeit	Relative Häufigkeit	
Wirtschaft/Technologie		24'038	0.12	22'997	0.10	-12.44
	Computing	3'217	0.02	1'037	0.00	-70.50
	Technologie/Wissenschaft	9'387	0.05	5'162	0.02	-49.67
	Wirtschaft	11'434	0.06	16'798	0.07	+34.46
Politik		15'541	0.08	18'171	0.08	+7.01
Gesellschaft/Geschichte/Militär		17'459	0.08	25'553	0.11	+33.96
	Militär/Konflikte	6'314	0.03	10'852	0.05	+57.31
	Geschichte/Gesellschaft	11'145	0.05	14'701	0.07	+20.73
Personen		22'758	0.11	49'814	0.22	+100.34
Gebaute/natürliche Umwelt		51'024	0.25	22'624	0.10	-59.42
Sport		28'729	0.14	35'134	0.16	+11.93
Kunst		37'806	0.18	42'317	0.19	+2.45
Bildung		9'219	0.04	9'091	0.04	-9.75

Die vorliegenden Resultate zeigen, dass die Akteure eine hohe Vielfalt aufweisen, wobei sehr unterschiedliche Aspekte unserer Gesellschaft vorkommen: von der Bildung über die Umwelt bis hin zur Geschichte. Die vertikale Organisation der Akteure zeigt sich als ein Mehrskalensystem. Die höchste Auflösung stellen dabei die einzelnen Akteure, die tiefste die acht Funktionen dar. Dazwischen gibt es neben den Subfunktionen die einzelnen Regionen mit thematischen Schwerpunkten. Je nach gestellter Fragestellung rückt eine unterschiedliche Ebene des Mehrskalensystems ins Zentrum der Betrachtung. Auf tiefster Auflösung lassen sich folgende Schlüsse ziehen:

- PERSONEN und SPORT sind sowohl absolut als auch relativ gesehen die zwei relevantesten Funktionen und sind durch eine hohe innere Homogenität gekennzeichnet.
- Höher, verglichen mit den anderen Funktionen ist die innere Heterogenität hingegen für die Funktionen WIRTSCHAFT/TECHNOLOGIE und GESELLSCHAFT/GESCHICHTE/MILITÄR. Erstens teilen sich diese Funktionen in drei (COMPUTING, TECHNOLOGIE/WISSENSCHAFT und WIRTSCHAFT) bzw. zwei (MILITÄR/KONFLIKTE und GESCHICHTE/GESELLSCHAFT) Subfunktionen auf. Zweitens umfassen sie ein breites Spektrum an latenten Themen (gelbe und hellgrüne Regionen in Abbildung 43). Die Relevanz, insbesondere die relative, ist für MILITÄR/KONFLIKTE (+57.31%), GESCHICHTE/GESELLSCHAFT (+20.73%) und

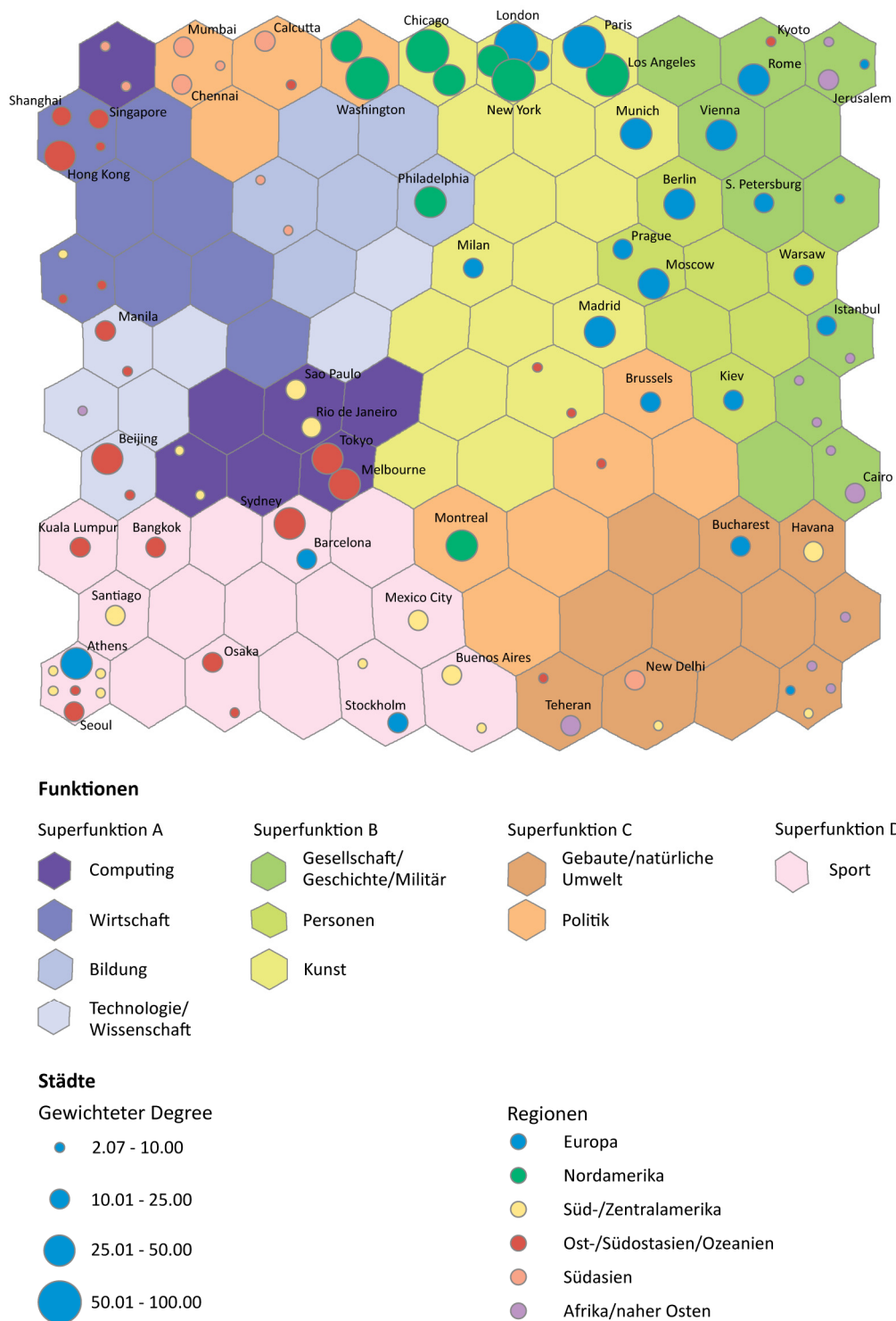
WIRTSCHAFT (+34.46%) hoch, für COMPUTING (-70.50%), sowie für TECHNOLOGIE/WISSENSCHAFT (-49.67%) tiefer.

- KUNST, BILDUNG und POLITIK nehmen sowohl hinsichtlich Relevanz als auch Heterogenität eine Zwischenstellung ein.
- Die Funktion GEBaute/NATÜRLICHE UMWELT ist trotz der hohen Häufigkeitswerte, relativ gesehen, wenig relevant.

7.3 Funktionale Städtenetze

Aus Kapitel 7.1 ergeben sich die Eigenschaften der Städte als Teil des globalen Städtenetzes: ihre lokale und globale Zentralität, die Rollen, die sie im Netzwerk übernehmen und ihre regionale Organisation. Die Perspektive hat sich anschliessend hin zu den Akteuren dieses Netzes und weg von den Städten gewandt (Kapitel 7.2). Daraus hat sich gezeigt, dass sich die Akteure auf acht Funktionen (WIRTSCHAFT/TECHNOLOGIE, POLITIK, GESELLSCHAFT/GESCHICHTE/MILITÄR, PERSONEN, GEBaute/NATÜRLICHE UMWELT, SPORT, KUNST und BILDUNG) zurückführen lassen. Zwei der Funktionen lassen sich zudem in zwei (MILITÄR/KONFLIKTE und GESCHICHTE/GESELLSCHAFT) bzw. drei (COMPUTING, TECHNOLOGIE/WISSENSCHAFT und WIRTSCHAFT) Subfunktionen unterteilen. Die folgenden Resultate führen die Hubs, und die Funktionen zu einem integrierten System zusammen. Abbildung 45 zeigt dieses integrierte Städte-Funktionen-System mittels eines SOM. Der Raum wird aufgrund des Anteils jeder Funktion und Subfunktion an der lokalen Zentralität der Hubs generiert.

Abbildung 45: SOM mit zwei Ebenen: Funktionen als hierarchisch organisierte Regionen und Städte als Punktsymbole



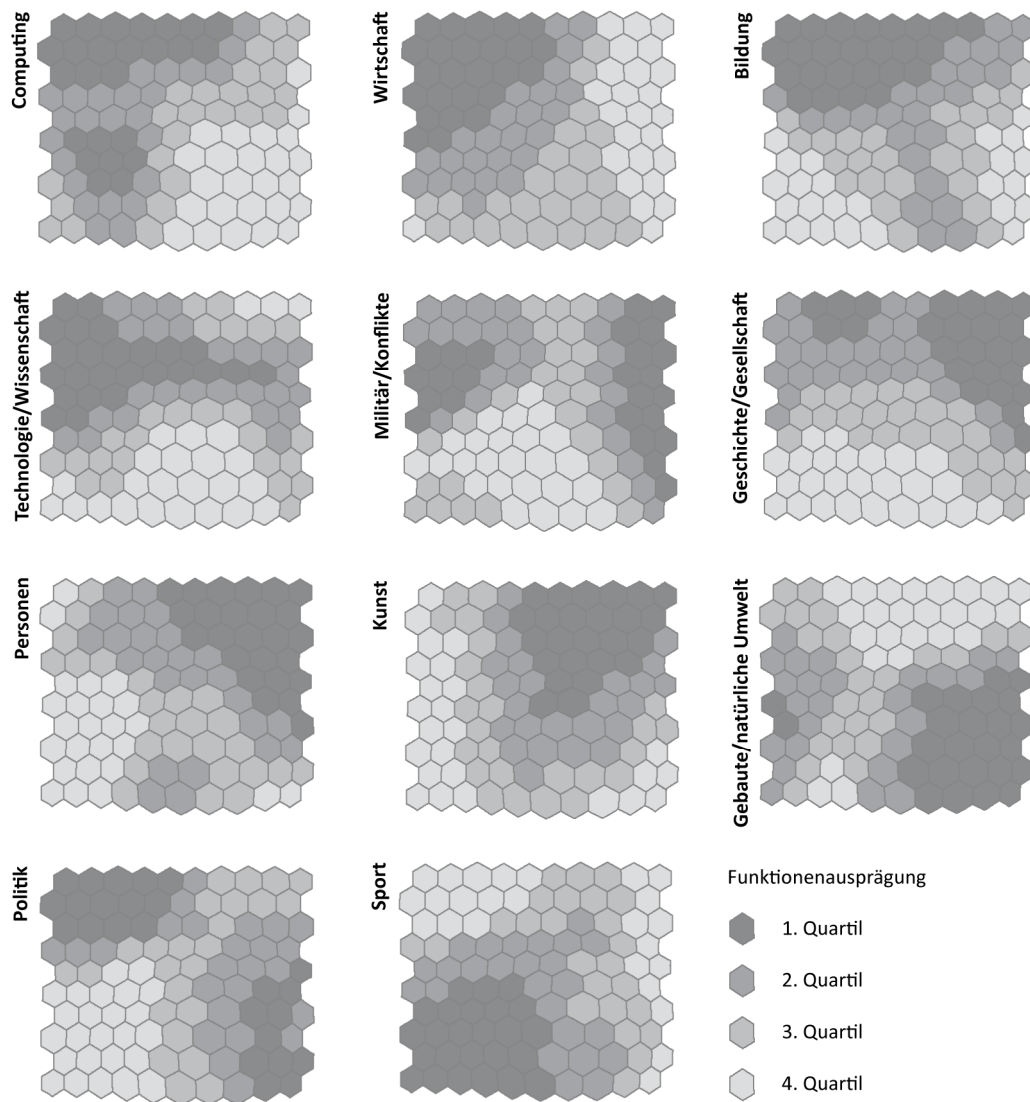
Die Neuronen geben die Funktionen und Subfunktionen als hierarchisches System mit drei Auflösungsstufen wieder: die 100 einzelnen Neuronen, zehn funktionale Regionen (durch Helligkeiten unterschieden) und vier (A-D) übergeordnete superfunktionale Regionen. Die zehn funktionalen Regionen und die vier superfunktionalen Regionen ergeben sich durch die Klassierung der Neuronen. Dabei werden von den elf ursprünglichen Funktionen und Subfunktionen GESCHICHTE/GESELLSCHAFT und MILITÄR/-

KONFLIKTE aufgrund ihres sehr ähnlichen Verlaufs im Raum gleich auf der tiefsten Aggregationsstufe zu einer Region zusammengefasst. Die drei Subfunktionen der Funktion WIRTSCHAFT/TECHNOLOGIE sind sich weniger ähnlich und werden nur auf der höheren Aggregationsstufe gruppiert und durch die Funktion BILDUNG ergänzt.

Jede funktionale Region ist somit massgeblich von einer bzw. zwei Funktionen oder Subfunktionen aus dem Vorkapitel bestimmt. Die unterschiedlichen Subfunktionen und Funktionen haben in der jeweiligen Region ihre stärkste Ausprägung. Die Regionen werden zusätzlich auch von weiteren Funktionen insbesondere aus benachbarten Regionen geprägt. Dies zeigt sich in Abbildung 46, wo der Verlauf im topologischen Raum der elf Funktionen und Subfunktionen durch Quartile dargestellt ist. Es wird auch ersichtlich, dass z.B. COMPUTING, WIRTSCHAFT, BILDUNG und TECHNOLOGIE/WISSENSCHAFT einen ähnlichen Verlauf im Raum haben, und dass die Neuronen mit der stärksten Funktionsausprägung teilweise überlappend sind. Diese Ähnlichkeit spiegelt die ähnliche Bedeutung dieser Funktionen für die einzelnen Städte wider. Konkret gilt der Grundsatz, dass COMPUTING, WIRTSCHAFT, BILDUNG und TECHNOLOGIE/WISSENSCHAFT für eine ähnliche Gruppe von Städten sehr bzw. weniger wichtig sind. Hingegen ist der Verlauf der Subfunktionen GESCHICHTE/GESELLSCHAFT und SPORT entgegengesetzt und somit auch deren Bedeutung für die Städte. Die Korrelationsanalyse zwischen den Anteilen der Funktionen pro Stadt bestätigt die Muster aus der SOM. Konkret ergeben sich folgende Strukturen:

- Die Korrelationswerte zwischen GESCHICHTE/GESELLSCHAFT, KUNST, PERSONEN und teilweise MILITÄR/KONFLIKTE sind mittel bis stark positiv (r zwischen 0.235 und 0.489). Diese vier Funktionen bilden eine gemeinsame, superfunktionale Region (SUPERFUNKTION B). Negativ ist der Zusammenhang insbesondere mit SPORT (r von -0.0297 bis -0.457) sowie GEBAUTER/NATÜRLICHER UMWELT (r von -0.064 bis -0.430).
- COMPUTING, WIRTSCHAFT, BILDUNG und TECHNOLOGIE/WISSENSCHAFT bilden eine gemeinsame, superfunktionale Region (SUPERFUNKTION A). Die positiven Korrelationswerte von TECHNOLOGIE/WISSENSCHAFT zu den anderen drei erwähnten Funktionen sind jedoch schwach bis sehr schwach (r von 0.005 bis 0.287). Negativ sind die Korrelationswerte insbesondere zu GEBAUTER/NATÜRLICHER UMWELT (r von -0.107 bis -0.294).
- POLITIK und GEBaute/NATÜRLICHE UMWELT bilden trotz schwachem Korrelationswert ($r=0.103$) eine gemeinsame superfunktionale Region (SUPERFUNKTION C), wobei sich diese auf zwei getrennte Teilbereiche aufteilt. GEBaute/NATÜRLICHE UMWELT weist negative Zusammenhänge zu allen übrigen Funktionen auf, POLITIK zeigt hingegen nur zu SPORT negative Werte.
- SPORT bildet eine eigenständige Region (SUPERFUNKTION D) und korreliert zu allen anderen Funktionen negativ.

Abbildung 46: Komponentenräume der elf Funktionen und Subfunktionen



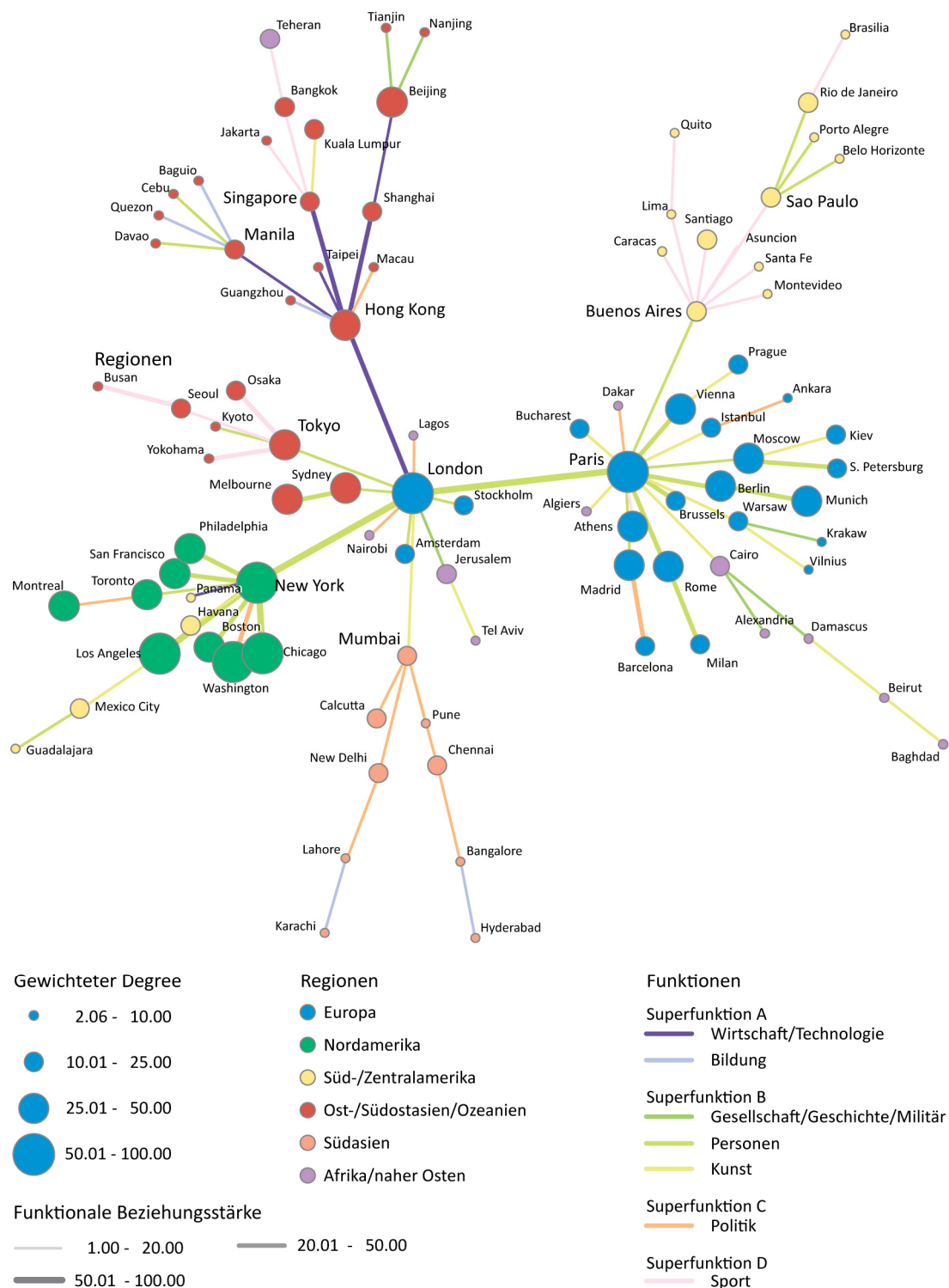
Die Punktsymbole (Abbildung 45) stellen die 95 Hubs dar. Die Grösse stellt die lokale Zentralität und die Farbtöne stellen die regionale Organisation dar. Je näher sich zwei Städte im funktionalen Raum positionieren, desto ähnlicher ist ihr funktionales Profil. So sind beispielsweise Tokyo und Melbourne funktional ähnlicher als Tokyo und Kyoto. Aus der gleichzeitigen Betrachtung der Funktionen und der Städte zeigt sich, dass zentralere Städte (z. B. Paris, London, New York) sich tendenziell im oberen Bereich des topologischen Raumes positionieren und somit weniger von den Funktionen SPORT und GEBaute/NATÜRLICHE UMWELT geprägt sind. Für wenig zentrale Städte (z.B. Stockholm, Buenos Aires, Teheran) sind in der Tendenz diese zwei Funktionen bestimmender. Dieses Muster wird teilweise von der Korrelationsanalyse bestätigt, so ist der Zusammenhang zwischen dem gewichteten *degree* und dem Anteil an Kunst ($r=0.566$) und an Personen ($r=0.335$) mässig bis stark und positiv. Der Zusammenhang zwischen dem gewichteten *degree* und dem Anteil an GEBAUTER/NATÜRLICHER UMWELT ($r=-0.460$) ist hingegen negativ.

Bei der Gegenüberstellung der Funktionen und der regionalen Organisation der Städte, fällt auf, dass EUROPÄISCHE und NORDAMERIKANISCHE Städte eher durch die Funktionen KUNST, PERSONEN und GESELLSCHAFT/GESCHICHTE/MILITÄR geprägt sind. Die afrikanischen und nahöstlichen Städte befinden sich im selben Bereich des topologischen Raumes. Bestimmend ist insbesondere die Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR. OST-/SÜDOSTASIEN/OZEANIEN haben hingegen in der Tendenz einen Schwerpunkt bei den Funktionen COMPUTING, WIRTSCHAFT, TECHNOLOGIE/WISSENSCHAFT und BILDUNG. Während sich SÜDASIEN insbesondere durch die POLITIK profiliert, weist SÜD-/ZENTRALAMERIKA ein heterogeneres Bild auf, wobei SPORT tendenziell eine überdurchschnittliche Bedeutung hat.

Generell greift die Zuweisung einer Stadt zu einer Funktion zu kurz und es kann somit nur die Rede von einer besonderen Bedeutung einer Funktion für eine Stadt sein, ohne zu vergessen, dass die lokale Einbettung einer Stadt durch alle Funktionen gegeben ist. Dies gilt insbesondere für die sehr zentralen Städte (London, Paris, New York, Washington, usw.), welche sich im topologischen Raum oben in der Mitte positionieren. Die vier Neuronen, welche diese Städte enthalten, weisen sehr hohe Ausprägungen sowohl der ÖKONOMISCH-TECHNOLOGISCHEN Funktionen (violett) als auch der SOZIO-KÜNSTLERISCHEN Funktionen (grün) auf.

Der multifunktionale Charakter des Netzwerkes wird in Abbildung 47 grafisch dargestellt. Die hervorgehobenen Informationen sind äquivalent zu der SOM in Abbildung 45. Die Städte und das MST-Netzwerk, welches diese Städte integriert, werden analog zu Abbildung 39 dargestellt: die Punktsymbole stellen die Städte, die Farbkodierung regionale Zugehörigkeit und die Grösse die Zentralität dar. Die Liniensymbole stellen jedoch nicht wie in Abbildung 39 die funktional undifferenzierten, interstädtischen Beziehungen dar, sondern die funktionalen Beziehungen mit der höchsten überdurchschnittlichen Stärke. Zu erwähnen ist, dass der MST nur 94 der 4'465 interstädtischen Beziehungen darstellt. Die aus der SOM gewonnen funktionalen Muster werden jedoch mehrheitlich bestätigt.

Abbildung 47: Multifunktionales globales Städtenetz zwischen den Hubs



Die Interaktionen zwischen den EUROPÄISCHEN und NORDAMERIKANISCHEN Städten sind, wie erwartet, überdurchschnittlich stark für die Funktionen PERSONEN, GESELLSCHAFT/GESCHICHTE/MILITÄR und KUNST. Die stärksten Beziehungen, z.B. Paris-London, London-New York und der NORDAMERIKANISCHEN Städte zu New York, gehören der Funktion PERSONEN an. Obwohl Athen in der SOM eine ausgeprägte SPORTLICHE Dimension aufweist, ist die Beziehung Athen-Paris überdurchschnittlich stark in

der politischen Funktion. Viele der nicht dargestellten Beziehungen von Athen zu den anderen Städten, insbesondere zu den Austragungsorten von olympischen Spielen, haben jedoch eine überdurchschnittliche Stärke in der Funktion SPORT. Sowohl der Schwerpunkt in der Funktion SPORT der SÜD-/ZENTRAL-AMERIKANISCHEN Region als auch der Schwerpunkt in der Funktion POLITIK der SÜDASIATISCHEN Region sind auch im Netzwerk ersichtlich. Die SÜD-/SÜDOSTASIATISCHE Region ist hingegen wie in der SOM in drei Bereiche unterteilt: erstens, die Beziehungen zwischen den zentralen Städten, mit Ausnahme von Tokyo, sind überdurchschnittlich stark in der Funktion WIRTSCHAFT/TECHNOLOGIE, zweitens, die Beziehungen zwischen den japanischen und koreanischen Städten sind von der Funktion SPORT geprägt, und drittens, die Beziehungen der peripheren Städte sind von unterschiedlichen Funktionen geprägt.

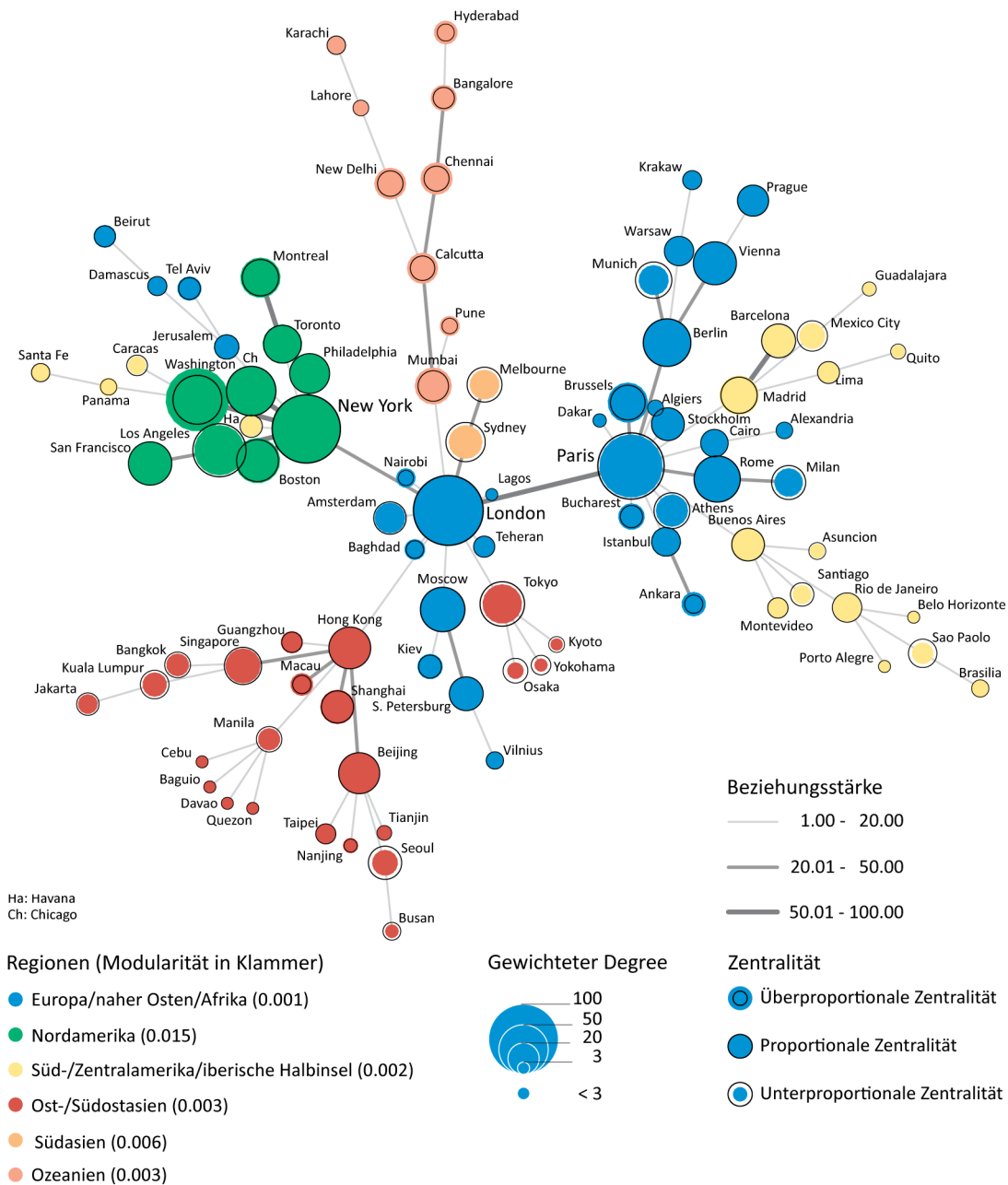
Die bisher beschriebenen Resultate beruhen ausschliesslich auf relativen Werten, sei es den Anteilen der Funktionen gemessen an der lokalen Zentralität der Städte oder dem Anteil der Stärke einer funktionalen Beziehung gemessen an der gesamten Beziehung. Bei den folgenden Erläuterungen rücken hingegen die absoluten Werte in den Vordergrund. Für zwei Funktionen, POLITIK und GESELLSCHAFT/GESCHICHTE/MILITÄR und einer Subfunktion, WIRTSCHAFT, werden die Netze zwischen den Hubs unter ausschliesslicher Berücksichtigung der Akteure, die zur jeweiligen Funktion gehören, berechnet. Die Funktionen und Subfunktionen gehören zu drei der vier übergeordneten Regionen in Abbildung 45. Die lokale Zentralität (gewichteter *degree*), die regionale Organisation der 95 Hubs und der MST dazwischen werden für die drei funktionalen Städtetze in den Abbildungen 45, 46 und 47 dargestellt.

POLITIK

Für die Funktion POLITIK erwartet man aufgrund der bisher vorgestellten Resultate, dass Brüssel, Montreal, Washington, Boston und vier südasiatische Städte eine zentralere Stellung im politischen Netzwerk als im allgemeinen Netzwerk innehaben. Dies wird in Abbildung 48 insbesondere für Washington bestätigt: der gewichtete *degree* im politischen Netzwerk (Grösse des grünen Kreises) ist überproportional verglichen mit dem gewichteten *degree* im allgemeinen Netzwerk (Grösse des Kreises mit schwarzer Kontur).

Bei einer detaillierten Betrachtung des politischen Netzes wird ersichtlich, dass die grobe Struktur derjenigen des allgemeinen Netzwerks sehr ähnlich ist. New York, London und Paris bilden weiterhin die Hauptzentren des Netzes. Die Verbindung London-New York verliert leicht an Stärke. Paris ist weiterhin das Zentrum Kontinentaleuropas und Südamerikas, London des ozeanischen und des asiatischen Raumes inklusive Russland und New York des nordamerikanischen Kontinentes und einer Minderheit der süd- und zentralamerikanischen Städte. Die afrikanischen Städte beziehen sich in gleicher Anzahl zu den drei Hauptzentren.

Abbildung 48: Globales Städtenetz zwischen den Hubs aufgrund der Funktion POLITIK



Auf regionaler Ebene ist die Auflösung der AFRIKANISCH/NAHÖSTLICHEN Region zu vermerken. Diese Städte sind nun Teil der EUROPÄISCHEN Region, die damit an Dichte einbüßt (Modularität = 0.001; Modularität für EUROPA im allgemeinen Netzwerk = 0.004). Die zwei ozeanischen Städte splittieren sich von der OST-/SÜDOSTASIATISCHEN Region ab und bilden eine eigenständige Region. Madrid und Barcelona gehören neu zur selben Region wie die süd- und zentralamerikanischen Städte. NORDAMERIKA bleibt in der Zusammensetzung unverändert, die Dichte nimmt jedoch zu (Modularität = 0.015; Modularität für NORDAMERIKA im allgemeinen Netzwerk = 0.008).

Auf der Ebene der einzelnen Städte zeigt sich ein Gewinn an lokaler Zentralität fast bei allen SÜDASIATISCHEN Städten, deren Stellung im MST bleibt jedoch nahezu unverändert. In der SÜD-/SÜDOSTASIATISCHEN Region bleiben Hong Kong, Tokyo und Peking

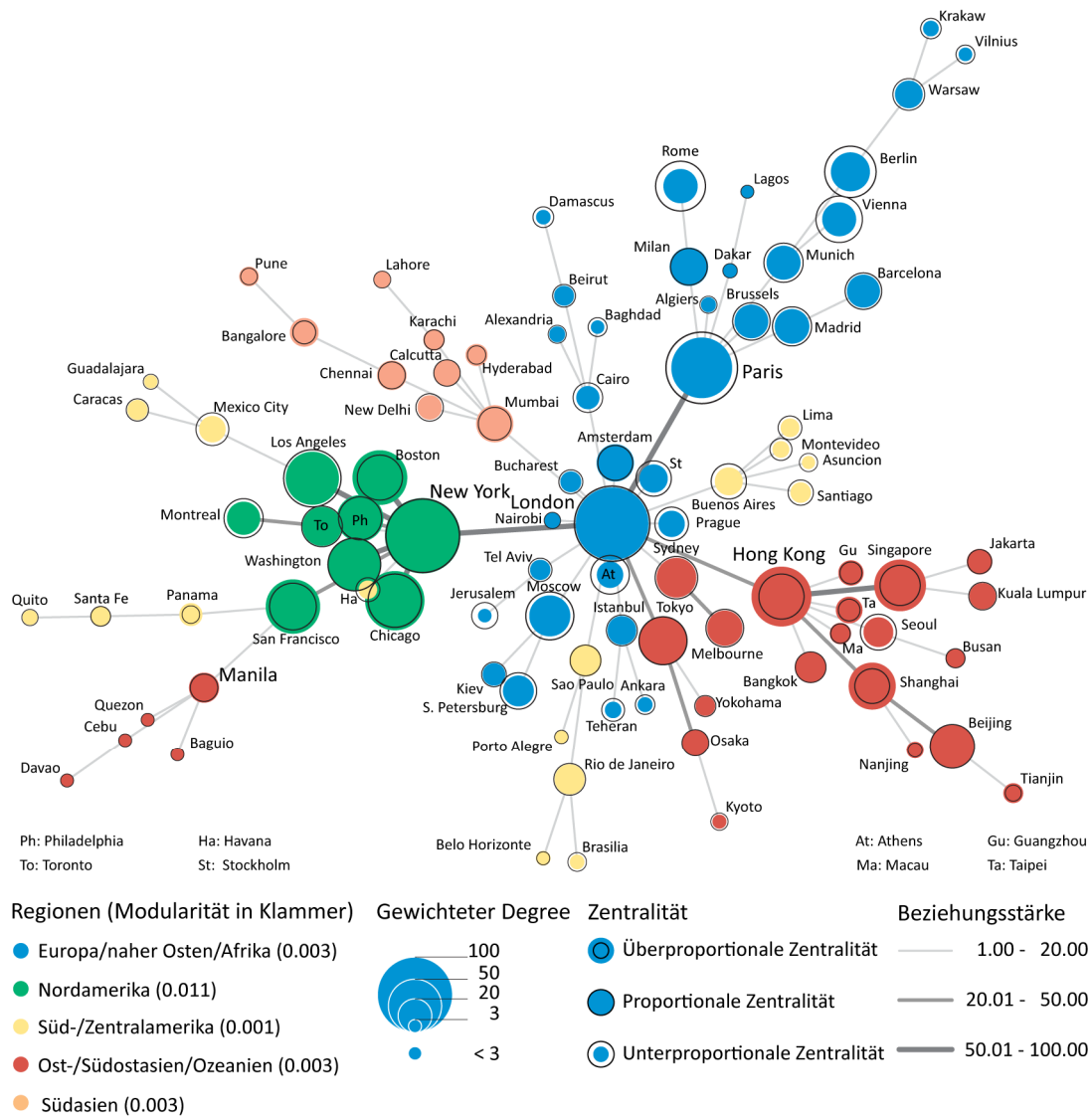
die wichtigsten Städte. Tokyo und alle weiteren Städte aus Japan und Korea verlieren jedoch an lokaler Zentralität. Die koreanischen Städte beziehen sich nun auf Peking und nicht mehr auf Tokyo. Hong Kong und Peking verstärken hingegen, trotz konstanter lokaler Zentralität, ihre Stellung im MST.

Die NORDAMERIKANISCHE Region sticht wegen ihrer Stabilität hervor. Washington rangiert nun aufgrund der lokalen Zentralität an dritter Stelle, also vor Paris, dennoch vermag es seine Stellung im MST nur gering zu verstärken. Nur Santa Fe, Panama und Karakas beziehen sich neu auf Washington. New York hält die sehr zentrale Stellung in der NORDAMERIKANISCHEN Region. Neu bezieht sich Jerusalem direkt und Tel Aviv, Damaskus sowie Beirut indirekt auf New York. Dies ist angesichts der engen politischen Beziehungen zwischen Israel und den USA nicht erstaunlich. Bei den EUROPÄISCHEN Städten verstärken insbesondere Berlin und London ihre Stellung ohne grössere lokale Zentralität. Andersherum bleibt die Stellung für Brüssel, Ankara und Bukarest trotz höherer lokaler Zentralität unverändert. Moskau bezieht sich neu auf London; Paris wird leicht geschwächt. Die nordafrikanischen Städte beziehen sich weiterhin auf Paris und die Städte der Sub-Sahara auf London. Wie erwähnt beziehen sich die nahöstlichen Städte, mit Ausnahme von Teheran und Bagdad, auf New York. In der SÜD-/ZENTRALAMERIKANISCHEN/IBERISCHEN HALBINSEL Region konkurriert Madrid neu mit Buenos Aires und profiliert sich als zweites Zentrum. Mexico City, Santiago und insbesondere Sao Paolo verlieren an Zentralität.

WIRTSCHAFT

Abbildung 49 stellt das Städtennetz zwischen den Hubs aufgrund der Subfunktion WIRTSCHAFT nach den gleichen Prinzipien wie Abbildung 48 für die Funktion POLITIK dar.

Abbildung 49: Globales Städtenetz zwischen den Hubs aufgrund der Subfunktion WIRTSCHAFT



Die aus Abbildung 45 erwartete Stärkung der Städte der Region OST-/SÜDOSTASIEN/OZEANIEN bei einer gleichzeitigen Schwächung der Region EUROPA/NAHER OSTEN/AFRIKA zeigt sich sehr deutlich. Hong Kong verstärkt sowohl die Stellung im MST als auch die lokale Zentralität und kristallisiert sich zum vierten Hauptpol im Netzwerk. Die Stärkung der OST-/SÜDOSTASIATISCH/OZEANISCHEN Städte zeigt sich auch an der höheren lokalen Zentralität von Shanghai, Singapur und Taipei und an den starken intraregionalen Beziehungen und der interregionalen Beziehung zwischen Hong Kong und London. Auf der anderen Seite schwächen sich in der Region EUROPA/NAHER OSTEN/AFRIKA sowohl die intraregionalen Beziehungen als auch die lokalen Zentralitäten der Städte, insbesondere von Rom, Berlin, Wien, Paris und Jerusalem ab. Paris ist auch in der Stellung im MST geschwächt. Während im allgemeinen Netzwerk eine Mehrheit der südamerikanischen, afrikanisch-nahöstlichen und europäischen Städte zum Pariser *hinterworld* gehören, bezieht sich nun eine Mehrheit davon auf London.

Auf regionaler Ebene ist wiederum die Auflösung der AFRIKANISCH-NAHÖSTLICHEN Region zu vermerken. Die SÜD-/ZENTRALAMERIKANISCHE Region ändert sich in der Zusammensetzung nicht, ändert jedoch die Organisation im MST und teilt sich nun in vier Gebiete auf: 1) Buenos Aires mit den weiteren südwestlichen Städten 2) Sao Paulo und die restlichen brasilianischen Städte 3) Panama und die nordwestlichen Städte 4) die zwei mexikanischen Städte und Karakas. Das Gebiet um Panama bezieht sich auf San Francisco, das um Mexiko City auf Los Angeles und die restlichen zwei auf London. Im allgemeinen Netzwerk beziehen sich die ersten zwei auf New York und die letzteren auf Paris.

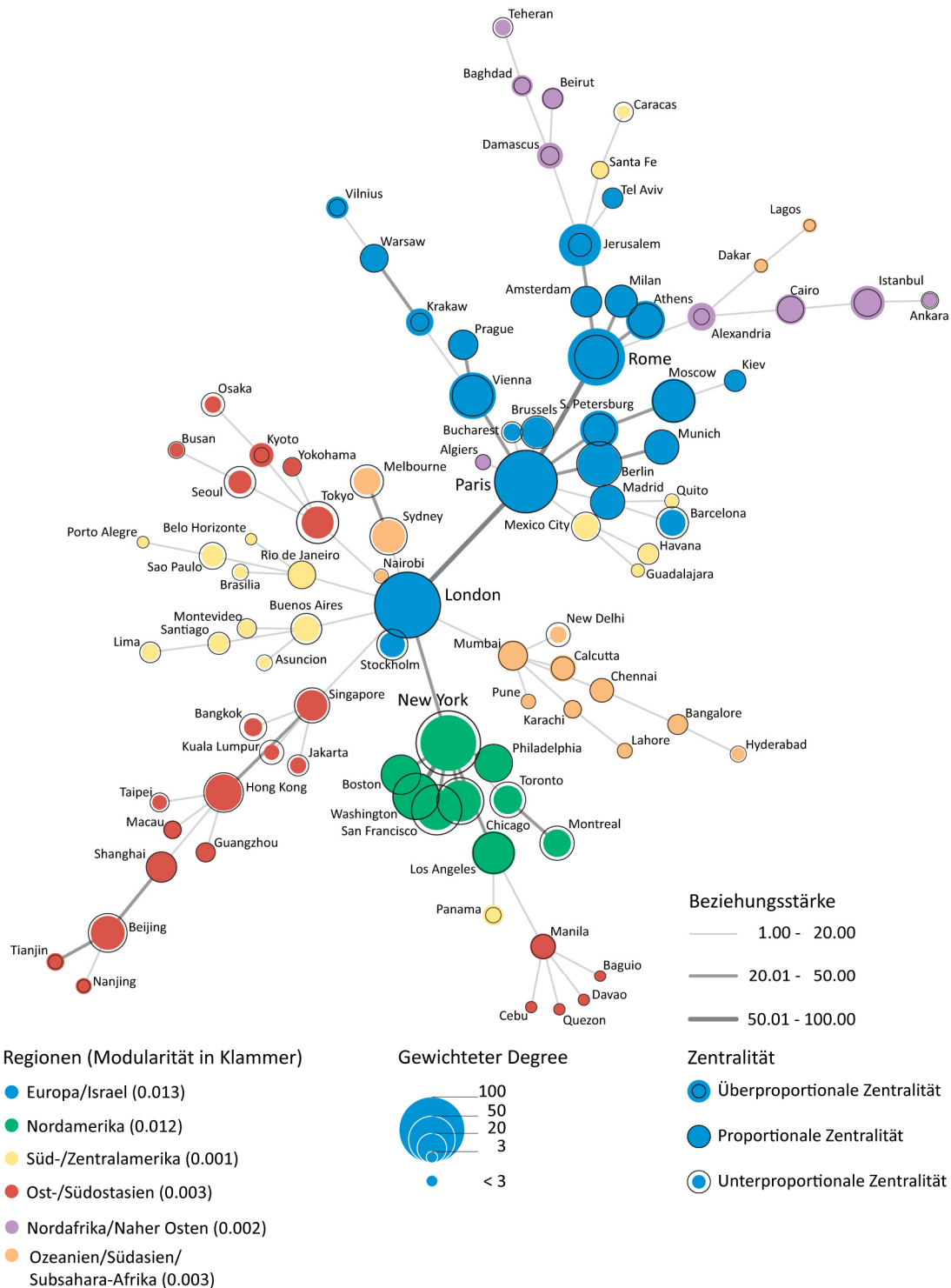
Die SÜDASIATISCHE und die NORDAMERIKANISCHE Region bleiben in der Zusammensetzung und der Organisation unverändert, jedoch wird in ökonomischer Hinsicht, neben Boston und Chicago, insbesondere San Francisco gestärkt. San Francisco ist die Hauptbezugsstadt für Manila und, wie oben erwähnt, Panama und hat eine höhere lokale Zentralität. Neben den philippinischen ändern auch die koreanischen Städte, wie bereits im politischen Netzwerk, ihre Stellung: sie beziehen sich nun auf Hong Kong und nicht mehr auf Tokyo.

Auf der Ebene des Gesamtnetzwerks ist die Erweiterung der Hauptachse New York-London-Paris um Hong Kong die prägendste Eigenschaft. Wobei New York und in noch grösserem Masse London erheblich zentraler als Hong Kong und Paris sind. London stärkt in wirtschaftlicher Hinsicht seine Rolle als zentralste Stadt des Systems.

GESELLSCHAFT/GESCHICHTE/MILITÄR

Die Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR positioniert sich im topologischen Raum (Abbildung 45) entfernt von der Funktion WIRTSCHAFT. Städte mit einem Schwerpunkt in dieser Funktion konzentrieren sich in Ost- und Zentraleuropa (Wien, Sankt Petersburg, Krakau und Vilnius) oder im mediterranen Raum (Rom, Jerusalem, Kairo, Istanbul, Damaskus, usw). Dies spiegelt sich auch im funktionalen Netzwerk (Abbildung 50) wider: die europäischen und arabischen Städte sind gestärkt, während insbesondere die OST-/SÜDOSTASIATISCHEN und teilweise die NORDAMERIKANISCHEN Städte an Zentralität verlieren. Die Ausnahme dabei ist Kyoto.

Abbildung 50: Globales Städtenetz zwischen den Hubs aufgrund der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR (Subfunktionen MILITÄR/KONFLIKTE und GESCHICHTE/-GESELLSCHAFT)



Unterschiedlich ist auch die regionale Organisation. Die EUROPÄISCHE Region erweitert sich um die zwei israelischen Städte. Die innere Dichte dieser Region ist mit einer Modularität von 0.013 (im allgemeinen Netzwerk Modularität = 0.004) überdurchschnittlich. Die AFRIKANISCH-NAHÖSTLICHE Region zerfällt in drei Teile. Die arabischen Städte bilden mit Istanbul und Ankara eine eigenständige Region, welche sich an der

Peripherie der Region EUROPA/ISRAEL positioniert. Die Städte der Sub-Sahara gruppieren sich neu mit den ozeanischen und südasiatischen Städten zu einer Region. Tel Aviv und Jerusalem gehören, wie erwähnt, zur Region EUROPA/ISRAEL. NORDAMERIKA, SÜD-/ZENTRALAMERIKA sowie OST-/SÜDOSTASIEN ändern die regionale Organisation nicht.

Auf individueller Ebene verstärken Rom, Jerusalem, Wien und Sankt Petersburg sowohl ihre lokale Zentralität als auch ihre Stellung im MST. Rom entwickelt sich zum Pol des mediterranen Raums, während das Pariser *hinterworld* mehrheitlich aus zentral- und osteuropäischen einerseits und iberischen und zentralamerikanischen Städten andererseits besteht. Die arabischen Städte, insbesondere Damaskus und Alexandria, sind in der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR zentraler. In der geschwächten OST-/SÜDOSTASIATISCHEN Region entwickelt sich ein Gleichgewicht zwischen Singapur, Hong Kong, Shanghai, Peking und Tokyo. Die Zentralität dieser fünf Städte ist etwa gleichwertig. Manila bezieht sich wie bereits im ökonomischen Netzwerk auf Los Angeles. Die NORDAMERIKANISCHEN Städte verlieren an Zentralität. Die Stellung im MST bleibt jedoch unverändert. Die ZENTRAL-/SÜDAMERIKANISCHE Region weist wiederum eine hohe Zersplitterung auf. Havanna und Guadalajara organisieren sich um Mexiko City, die südwestlichen Städte um Buenos Aires, die brasilianischen um Rio de Janeiro. Panama bezieht sich auf Los Angeles, Santa Fe und Karakas auf Jerusalem und Quito auf Madrid.

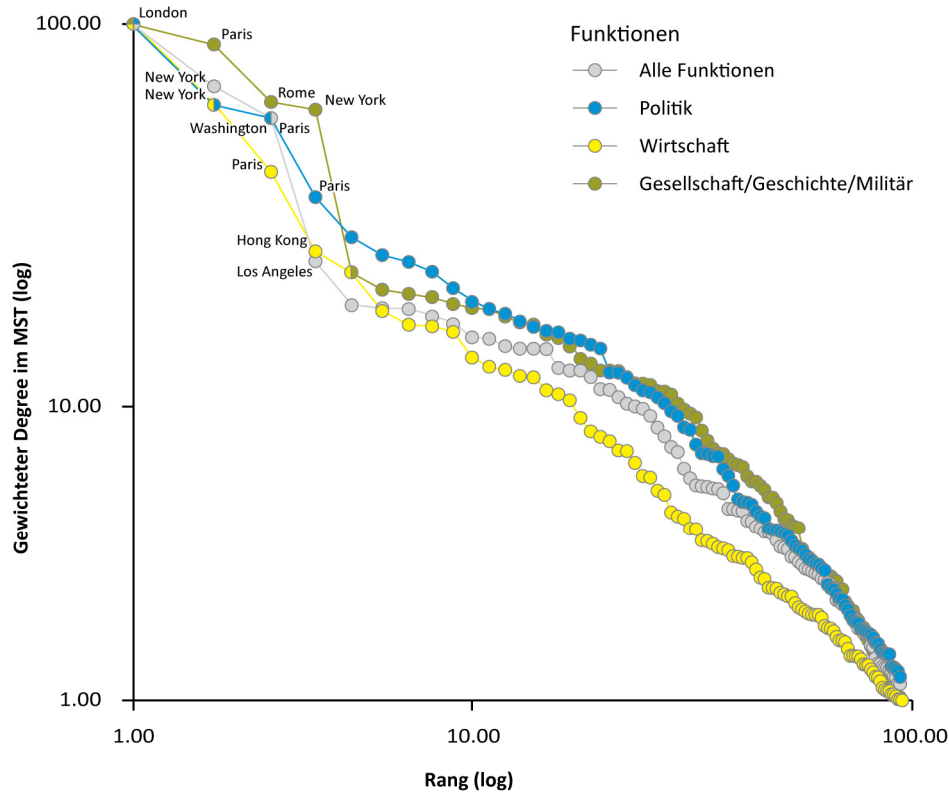
Auf der Ebene des Gesamtnetzwerkes ist die hohe Modularität, die auf überdurchschnittlich starke intraregionale Verbindungen hindeutet, zu vermerken. Die Modularität beträgt 0.554 und ist somit fast doppelt so hoch wie im allgemeinen und den anderen beiden funktionalen Netzwerken. Der MST besteht wie das ÖKONOMISCHE Netzwerk aus vier Hauptpolen. Der zusätzliche Pol, nebst London, Paris, und New York, ist jedoch Rom. Die vier Städte sind nicht wie in der Funktion WIRTSCHAFT sternförmig um London herum angeordnet, sondern organisieren sich entlang der Hauptachse des Netzwerks. London und Paris bilden mit gleichwertiger Zentralität (London = 100, Paris = 93) die Mitte dieser Achse. Rom (gewichteter *degree* = 75) und New York (gewichteter *degree* = 71) sind etwas peripherer gelegen.

Fazit

Die Betrachtung der drei funktionalen Netzwerke und des allgemeinen Netzwerks ergibt ein dichotomes Bild. Einerseits bleiben grundlegende Strukturen konstant. Dies zeigt sich in Abbildung 51, welche die Verteilung der Zentralitäten in den drei funktionalen Netzwerken und im allgemeinen Netzwerk vergleicht. Der Verlauf der vier Kurven ist tendenziell gleich. Insbesondere die Zentralitäten der peripheren Städte sind fast deckungsgleich. Andererseits gibt es klare und prägende Unterschiede, insbesondere was die Spitze der Hierarchie angeht. So sind das allgemeine und das POLITISCHE Netzwerk tendenziell dreipolig, und das WIRTSCHAFTLICHE sowie das Netzwerk der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR aber vierpolig. Im POLITISCHEN Netzwerk besteht der nordamerikanische Pol allerdings aus zwei Städten, New York und Washington. Im

WIRTSCHAFTLICHEN Netzwerk ist die Abnahme der Zentralitäten der vier Hauptpole stark exponentiell.

Abbildung 51: Log-Log-Diagramm des gewichteten *degree* im MST die drei funktionale Netzwerke Funktionen und das allgemeinen Städtenetz



Auf regionaler Ebene sind sowohl Ähnlichkeiten als auch Unterschiede festzustellen. So bleibt die regionale Organisation in den Grundsätzen gleich, kleinere Unterschiede sind hingegen funktionsabhängig. Die EUROPÄISCHE Region ändert je nach Funktion ihre Zusammensetzung, bleibt aber als Region stets sehr zentral. Die Zentralität ist im WIRTSCHAFTLICHEN Netzwerk am schwächsten. Paris und London sind immer die zentralsten Städte. Änderungen in der regionalen Hierarchie gibt es trotzdem: Berlin und Brüssel verstärken sich im politischen Netzwerk, Rom und Wien im SOZIO-HISTORISCH-MILITÄRISCHEN Netzwerk.

NORDAMERIKA bleibt unverändert in der Zusammensetzung und im hohen Anteil an intraregionalen Beziehungen. New York ist das Zentrum der Region. Washington schliesst im POLITISCHEN Netzwerk zu New York auf.

OST-/SÜDOSTASIEN variiert in der Zentralität sehr stark. Die Region ist in ökonomischer Hinsicht vergleichbar zentral wie EUROPA und NORDAMERIKA, hingegen peripherer aus SOZIO-HISTORISCH-MILITÄRISCHER Perspektive. Hong Kong, Tokyo, Peking, Singapur und Shanghai geben der Region den plurizentrischen Charakter.

Die Stellung von Sydney und Melbourne ist im MST unverändert. Die beiden Städte bilden immer einen eigenen Ast und beziehen sich auf London. Die Regionszugehörigkeit ist funktionsabhängig.

Die SÜD-/ZENTRALAMERIKANISCHE Region ist von einer hohen Fragmentierung und geringen Zentralität geprägt. Sie bezieht sich sowohl auf Europa (Paris, London und Madrid) als auch auf Nordamerika (New York, Los Angeles, San Francisco und Washington). Die Zusammensetzung bleibt unverändert mit Ausnahme vom POLITISCHEN Netzwerk, in dem Madrid und Barcelona Teil der Region sind. Buenos Aires, Mexiko City und abwechselnd Sao Paolo und Rio de Janeiro bilden die Zentren der Region.

SÜDOSTASIEN ist mit Ausnahme von der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR eigenständig und unverändert. Die Stellung im Netzwerk ist stets peripher und mit der Hauptbeziehung zu London. Der funktionale Schwerpunkt der Region ist die POLITIK.

Die AFRIKANISCHEN/NAHÖSTLICHEN Städte sind stets peripher und bilden für die drei betrachteten Funktionen nie eine eigenständige Region. Aus einer WIRTSCHAFTLICHEN und POLITISCHEN Perspektive sind sie Teil Europas, im SOZIO-HISTORISCH-MILITÄRISCHEN Netzwerk gehören sie drei unterschiedlichen Regionen an.

8 Diskussion

Im diesem Kapitel werden die Ergebnisse in Bezug auf die Forschungsfragen und den dargelegten Stand der Forschung aufgezeigt und diskutiert. Der erste Teil befasst sich mit der ersten Forschungsfrage und erörtert den Beitrag von Wikipedia als Datengrundlage im Forschungsbereich der *global cities* und *world cities*. Im zweiten Teil werden die vorgeschlagenen methodischen Ansätze kritisch diskutiert. Der letzte Teil befasst sich mit dem thematischen Aspekt und setzt die Resultate aus den drei Fallstudien in Bezug zu den theoretischen und empirischen Erkenntnissen aus der Literatur.

8.1 Datengrundlage

Die von der Globalisierung und dem technologischen Wandel angetriebenen Urbanisierungsprozesse, insbesondere die Integration der Städte in ein globales Städtetz, sind hoch komplex. Dies stellt besondere Ansprüche an die Datengrundlage, auf die sich die Analyse und Exploration dieser Prozesse stützt. Die etablierten Datengrundlagen, die Intrafirmenbeziehungen von APS (Taylor 1997; Taylor et al. 2002a, 2002b), die Besitzverhältnisse in transnationalen Unternehmen (Wall und Van der Knaap 2009; Rozenblat 2004) oder das Flugpassagiertvolumen (Smith und Timberlake 1995a, 2001) decken diese Ansprüche nur teilweise ab. Die Analysen dieser Arbeit greifen auf Wikipedia als Datengrundlage zurück, mit dem Ziel, ihre Eignung zu evaluieren und somit dieses Defizit auszugleichen. Die vorliegende Diskussion konfrontiert Wikipedia mit Anforderungen aus der Stadtgeographie und vergleicht sie mit den erwähnten, etablierten Datengrundlagen. Tabelle 25 fasst die betrachteten Eigenschaften der Datengrundlagen zusammen.

Im zweiten Teil dieses Kapitels wird der Einfluss der Risiken bei der Verwendung von Wikipedia auf die vorliegenden Analysen evaluiert.

Tabelle 25: Vergleich der Eigenschaften von Wikipedia mit den etablierten Datengrundlagen.

	Intrafirmen- beziehungen APS	Besitz- verhältnisse transnat. Unter- nehmen	Flugpassagier- volumen	Wikipedia
Relationale Struktur	Angenommene Intrafirmenbezie- hungen	Besitzverhältnisse	Flugbewegungen	Linkstruktur im Artikelsystem
Dualität	APS-Stadt	Unternehmen- Stadt	-	Akteur-Stadt
Globale Abdeckung	316 Städte	6'308 Städte	ca. 100 Städte	3'532 Städte
Multifunktionali- tät	ökonomisch	ökonomisch	allgemein	8 Funktionen bzw. 11 Subfunktionen
Zeit	Moment- aufnahme	Moment- aufnahme	Moment- aufnahme	Zeitliche Tiefe

Aus der Theorie zu den globalen Städten kommt insbesondere die relationale Natur (Smith und Timberlake 2001) des Integrationsprozesses hervor. „Globales Netzwerk an Städten“ (Friedmann und Wolff 1982; Sassen 2002; Taylor 2004) und „Ströme“ (Castells 2001) sind dabei nur zwei der Begriffe, die dieses Charakteristikum unterstreichen. Von den etablierten Datengrundlagen sind insbesondere die Besitzverhältnisse in transnationalen Unternehmen (Rozenblat 2004; Wall und Van der Knaap 2009) oder das Flugpassagiervolumen (Smith und Timberlake 1995a, 2001) explizit relational. Die Datengrundlagen mit den Geschäftsniederlassungen der APS (Taylor 1997; Taylor et al. 2002a, 2002b) sind per se nicht relational. Die relationale Komponente wird aufgrund von wenig wahrscheinlichen (Rozenblat 2004) Annahmen modelliert.

Die aus Wikipedia abgeleitete Datengrundlage zur Analyse des globalen Städtensetzes ist explizit relational, weil sie sich auf das Artikelsystem und somit auf die Links zwischen den Artikeln stützt. Artikel, welche Städte beschreiben, stellen die Städte dar. Die Beziehungen stützen sich auf ungefähr fünf Millionen explizite semantische Verbindungen zwischen den Artikeln. Die Interaktionsstärke zwischen den Städten wird aufgrund der Anzahl und der Spezifität der Verlinkungen bestimmt. Konkret werden die Beziehungen aufgrund der in den Städte-Artikeln einkommenden Links aus gemeinsamen Artikeln (Abbildung 13) modelliert. Damit gibt der Datensatz auch den dualen Charakter als weitere grundlegende Eigenschaft des globalen Städtensetzes wieder. Die Dualität ergibt sich aus der Wechselwirkung zwischen den Städten als räumlichen Ankerpunkten und den urbanen Akteuren, welche durch ihr Handeln das Netzwerk aufspannen (Pumain 2006a; Pflieger und Rozenblat 2010). Die oben erwähnten gemeinsamen Artikel, Akteur-Artikel genannt, entsprechen den urbanen Akteuren, die Städte-Artikel den Städten.

Neben dem relationalen und dem dualen Aspekt ist zudem die Globalität der Datengrundlage (Short et al. 1996; Beaverstock et al. 2000b) entscheidend. Der erarbeitete Datensatz umfasst 3'532 Städte. Dies entspricht den Städten, die nach den Informationen in Wikipedia mehr als 100'000 EinwohnerInnen haben. Die Anzahl ist geringer verglichen mit dem Datensatz von 6'308 Städten von Alderson und Beckfield (2004) aber erheblich grösser im Vergleich zu Taylor et al. (2002b) (316 Städte). Das Potential von Wikipedia wird in dieser Arbeit jedoch bei Weitem nicht ausgeschöpft. Es gibt in der englischsprachigen Wikipedia etwa 150'000 Artikel, die mit einer Kategorie vermerkt sind, die das Wort „*cities*“ oder „*city*“ enthalten. Auch wenn man annimmt, dass die Hälfte davon keine realen Städte, sondern Artikel mit einem städtischen Bezug (z.B. „*subway*“ oder „*atlantis*“) sind, und dass nicht alle Artikel genügend verlinkt sind, um daraus robuste Aussagen abzuleiten, ist die Anzahl immer noch erheblich. In dieser Hinsicht ist eine umfassende Exploration der Struktur der Städtetzwerke nicht nur auf globaler Ebene, wie in dieser Arbeit, sondern auch mit einer feineren Granularität, d.h. auf kontinentaler, nationaler und teilweise sogar regionaler Ebene möglich.

Der Schlüsselbeitrag dieser Arbeit für die stadtgeographische Forschung liegt weder bei der relationalen noch bei der globalen Dimension der Datenbasis, obwohl auch diesen beiden Aspekten eine angemessene Beachtung entgegen gebracht wird. Das Kernelement ist die Multifunktionalität der Datengrundlage und die Möglichkeit, Funktionen zu unterscheiden. Die bisherigen multifunktionalen empirischen Arbeiten beschränken sich auf einen Artikel von Taylor (2005a), obwohl nach Alderson und Beckfield (2004) ein multifunktionaler Datensatz als „ideal“ beschrieben wird. Mit dem vorliegenden Datensatz wird dieses Ideal erreicht. Jeder Akteur bzw. jeder Akteur-Artikel in Wikipedia wird aufgrund des Artikelstitels und der Kategorien, die dessen Inhalt widerspiegeln, einer Funktion zugewiesen. Die Definition der Funktionen erfolgt sowohl mit etablierten semantischen Analysen als auch mit den im halbstrukturierten Kategoriensystem von Wikipedia enthaltenen holonymischen und hyperonymischen Beziehungen. Der Detaillierungsgrad der Funktionsdefinition und die Anzahl Funktionen ist nicht a priori gegeben, sondern ergibt sich aus den Daten. Der Detaillierungsgrad lässt sich beliebig, der Fragestellung entsprechend, bestimmen und variiert zwischen jedem einzelnen der um die Million Akteure auf der tiefsten Aggregationsstufe und den acht hervorgehobenen Funktionen auf der höchsten Aggregationsstufe. Zum Beispiel ist es im politischen Themenbereich möglich, die Gesamtheit der politischen Akteure, nur die Themen mit Bezug zu Regierungen und Wahlen, oder auf der tiefsten Aggregationsebene einzelne Wahlen (z.B. Präsidentschaftswahl der USA), zu betrachten. Diese multiskalare Differenzierung der Funktionen stellt einen erheblichen Mehrwert von Wikipedia im Vergleich zu den etablierten Datengrundlagen dar, weil damit gezieltere und präzisere Fragestellungen beantwortet werden können. Zudem ist nicht nur das funktionale Spektrum breiter und differenzierter, sondern auch in quantitativer Hinsicht setzt Wikipedia neue Massstäbe. Zum Vergleich werden in der vorliegenden Arbeit etwas weniger als eine halbe Million urbane Akteure (im Durchschnitt knapp 120 pro Stadt) betrachtet, während z.B. Taylor (2004) 7'500 Niederlassungen von APS (im Durchschnitt 23 pro

Stadt) und Rozenblat (2004) 4'000 Tochtergesellschaften (im Durchschnitt 10 pro Stadt) in ihre Analysen mit einbeziehen.

Ein weiterer Unterschied ist die zeitliche Dimension der betrachteten Akteure. Die oben erwähnten Modellierungsansätze sind Momentaufnahmen. Sie betrachten einen kurzen, genau eingegrenzten zeitlichen Abschnitt. Falls vor diesem Zeitpunkt eine Niederlassung (Taylor 1997; Taylor et al. 2002a, 2002b) geschlossen, oder eine Tochtergesellschaft (Rozenblat 2004; Wall und Van der Knaap 2009) verkauft, oder noch eine Flugstrecke (Smith und Timberlake 1995a, 2001) eingestellt wird, werden diese in keiner Form in den Analysen berücksichtigt. Gleichermassen werden für die Zukunft geplante Änderungen auch nicht mit einbezogen. In Wikipedia ist der zeitliche Aspekt komplexer und vielfältiger. Wichtige Ereignisse der Vergangenheit und Zukunft werden ebenso wie das Gegenwärtige berücksichtigt. In dieser Hinsicht wird sowohl die Zentralität von Berlin vom Austragen der olympischen Spiele 1936, als auch diejenige von Vancouver vom Austragen der Spiele in 2010, sowie diejenige von Sochi von den olympischen Spielen 2014 beeinflusst. Die zeitliche Dimension ist nicht nur von den Ereignissen der Gegenwart bestimmt, sondern auch von der Einschätzung, die man in der Gegenwart von Ereignissen der Vergangenheit und der Zukunft vornimmt. Wikipedia wurde 2001 lanciert (Wikipedia contributors 2011d) und verbucht seither einen stetigen Zuwachs an Einträgen. Zudem haben Flanagin und Metzger (2008) festgestellt, dass zeitlich nahe Ereignisse eine höhere Resonanz in verfassten Artikeln finden. Dies führt dazu, dass die Inhalte eine zeitliche Verzerrung aufweisen. Diese beeinflusst die Validität der Resultate durch das Prinzip der grossen Zahlen insgesamt nicht, kann jedoch bei Analysen nicht gänzlich ignoriert werden.

Es ist extrem komplex einzuschätzen, welche zeitliche Dimension die urbane Dynamik besser widerspiegelt. Ein kleines Beispiel aus Zürich soll jedoch veranschaulichen, dass die zeitliche Dimension in Wikipedia aus stadtgeographischer Perspektive zumindest plausibel ist. Swissair hat die Schweiz und Zürich massgeblich geprägt und hinterlässt auch nach dem Konkurs ein Erbe für die Stadt Zürich, welches umfassender als die neu gegründete Swiss ist. Swiss beeinflusst insbesondere die Interaktionen erster Ordnung von Zürich mit anderen Städten. Die Erinnerung und das Image der Swissair färben jedoch im Positiven und im Negativen noch auf Zürich ab und beeinflussen somit zumindest die Interaktionen zweiter Ordnung. Genau dieses Element wird auch in Wikipedia aufgenommen.

Dank der Globalität, der Relationalität, der zeitlichen Tiefe und insbesondere der funktionalen Breite und der multiskalaren Differenzierung übertrifft Wikipedia die etablierten Datengrundlagen. Wikipedia birgt trotz allen Vorteilen in Bezug auf die Modellierung eines globalen Städtetetzes auch Risiken: einerseits zeigen sich räumliche Verzerrungen der Inhalte bzw. ein inhaltliches angelsächsisches Übergewicht in der englischsprachigen Wikipedia (Flanagin und Metzger 2008) und andererseits besteht eine gewisse Volatilität (die sich von der laufenden Aktualisierung ergibt) sowie Ungenauigkeit der Inhalte, insbesondere der Links (Denning et al. 2005).

Die Evaluation der Datengrundlage hat ergeben, dass die erwähnten Risiken die Resultate nicht massgeblich beeinflussen oder wenigstens eingeschätzt und korrigiert werden können. Die Risiken werden insbesondere durch den Modellierungsansatz reduziert. Ein Stadt-Artikel hat im Durchschnitt achtmal mehr einkommende (ca. 930 Links) als ausgehende Links (ca. 120 Links). Bei den zentralen Städten ist das Verhältnis noch ausgeprägter. Dadurch ist die Datengrundlage nicht aufgrund der grossen Zahlen risikoarmer, sondern auch wegen der Differenziertheit des Inhalts. Nach Flanagin und Metzger (2008) führt eine höhere Anzahl AutorInnen zu höherer Genauigkeit und Glaubwürdigkeit der Einträge. Man kann zusätzlich die Annahmen treffen, dass eine grössere Anzahl betrachteter Artikel auch einer höheren Anzahl AutorInnen und somit einer höheren Genauigkeit und Glaubwürdigkeit entspricht. Da sich die einkommenden Links auf den Inhalt von etwa einer Million Akteur-Artikeln und nicht nur auf den von 3'532 Stadt-Artikeln stützen, sollten die Glaubwürdigkeit und die Genauigkeit höher sein. Trotzdem werden in der vorliegenden Arbeit die Volatilität der Einträge und die räumliche Verzerrung explizit untersucht. Der Einfluss der Volatilität und der Genauigkeit der Einträge werden mit dem Auswählen von 75% der Akteur-Artikel und dem entsprechenden Vernachlässigen von 25% derselben getestet. Der Einfluss ist minim und die Resultate sind somit hinsichtlich der Änderungen der Linkstruktur sehr robust.

Die räumliche Verzerrung spiegelt sich in einem leichten Übergewicht der US-amerikanischen Städte wider. Dieses ist bei einfachen netzwerkanalytischen Massen, wie dem gewichteten *degree* stärker als bei komplexeren Massen wie dem Z-Wert. Insgesamt wird die räumliche Verzerrung als gering und unproblematisch eingeschätzt. Die Berechnung des MST hat sich als eine effektive Methode zur Kompensation der räumlichen Verzerrungen erwiesen. Die Verzerrungen reduzieren sich damit noch weiter und die grundlegenden Strukturen des Städtenetzes werden damit hervorgehoben. Positiv auf die Minderung der räumlichen Verzerrungen wirkt sich auch die Hubdefinition nach Guimerà und Amaral (2005a) aus. Während beim gewichteten *degree* die US-amerikanischen Städte eine überdurchschnittliche Zentralität haben, ist bei der Hubdefinition, insbesondere beim Z-Wert, keine räumliche Verzerrung mehr erkennbar.

Einen Bezug zum Raum und der Genauigkeit der Links hat die Mehrdeutigkeit einzelner Toponyme. Dieser Punkt wird zwar nicht systematisch evaluiert, aus den Resultaten lässt sich jedoch bei einzelnen, peripheren Städten vermuten, dass die AutorInnen die Links aufgrund der Mehrdeutigkeit der Toponyme der Städte falsch gesetzt haben. Ein explizites Beispiel dafür ist Cartagena. Das Toponym bezeichnet drei unterschiedliche Städte in Spanien, Kolumbien und Chile. Der Einfluss ist für die relationalen Eigenschaften dieser Städte erheblich, gleichwohl ist der Einfluss auf das Gesamtnetz gering und unbedeutend.

Ein letzter entscheidender Punkt ist die Validität der Resultate. Nur wenn die gewonnen Erkenntnisse plausibel und wenn sie den erwarteten Eigenschaften der untersuchten Urbanisierungsprozesse entsprechen, ist die Datengrundlage für die Analyse geeignet. Die thematischen Resultate werden im Einzelnen im letzten Unterkapitel diskutiert,

jedoch kann die Validität der Resultate und somit der Datengrundlage an dieser Stelle vorweggenommen werden. Die Strukturen des globalen Städtetetzes, die Eigenschaften der Städte, die eruierten Funktionen, sowie die dargelegten funktionalen Städtetetze scheinen valid und plausibel. Eine detaillierte Diskussion folgt im letzten Unterkapitel. Ein weiterer Punkt, der hinsichtlich der Validität hervorzuheben ist, ist intrinsisch in Wikipedia vorhanden: Wikipedia ist als *crowdsourcing*-Projekt ein Produkt der Informationsgesellschaft und kommt somit, wie das Städtetetz, aus der Informationsgesellschaft hervor. Der gemeinsame Ursprung, die Einbettung der Daten und des Untersuchungsgegenstandes im gleichen Zeitalter und –geist können als eine Verstärkung der Validität der Datengrundlage verstanden werden.

Zusammenfassend lässt sich die erste Forschungsfrage mit ja beantworten: Wikipedia ist eine Antwort auf das „*dirty little secret*“ (Short et al. 1996) und somit eine valide und robuste Datengrundlage für die Erforschung des globalen Städtetetzes und vermutlich auch der kontinentalen, nationalen und regionalen Städtetetze. Dieser Schluss befreit zukünftige Forschende trotzdem nicht von der systematischen Validierung der sich stetig wandelnden Datengrundlage und öffnet den Weg für weiterführende Untersuchungen basierend auf semantische Daten.

Dabei ergeben sich zwei Stossrichtung: einerseits die Verfeinerung der Modellierung mithilfe von Wikipedia und andererseits die Erschliessung weiterer unstrukturierter semantischer Datengrundlagen. In beiden Fällen bietet der Forschungsbereich des *natural language processings* (NLP) umfassende methodische Ansätze. *Geographical information retrieval* ist Teil des NLP-Forschungsbereichs und widmet sich der Erkennung und Strukturierung von geographischen Informationen in Texten (Jones und Purves 2008). Folgt man den methodischen Ansätzen dieser Forschungsrichtung, so können Nennungen von Städten, die in den Akteur-Artikeln nicht explizit mit den jeweiligen Stadt-Artikeln verbunden sind (fehlende Links) erkannt werden. Zudem können die aufgrund der Mehrdeutigkeit der Toponyme falsch gesetzten Links erkannt werden. Beide Ansätze verbessern die Qualität der Datengrundlage.

Die Stärken der Methoden des NLP liegen jedoch nicht in der Ergänzung der vordefinierten Strukturen von Wikipedia, sondern in der Erschliessung unstrukturierter Datengrundlagen. Die Untersuchung der strukturellen Elemente von Wikipedia reduziert die Risiken der wissenschaftlichen Nutzung dieses enzyklopädischen Wissens (Hecht 2007). Gleichwohl führen die vordefinierten Strukturen zu Verzerrungen in den Resultaten, wie bereits oben erwähnt. Die Erschliessung von unstrukturierten Daten vermeidet diesen strukturellen Determinismus. Umsetzungsbeispiele dafür sind: die Modellierung des Städtetetzes aufgrund der Kookkurrenz von Städtenamen in Zeitungsartikeln, in ausgewählten Blogs oder in funktionsspezifischen Texten (z.B. Geschäftsberichte im ökonomischen Bereich). Mit der Erkennung des Themas, mit dem sich der Textausschnitt, in dem die beiden Stadtnamen vorkommen, befasst, können die Funktionen identifiziert werden. Als Methode zur Erkennung des Themas kann das *topic modeling* (Steyvers und Griffiths 2007) dienen. Das *opinion mining* befasst sich hingegen mit der Erkennung und

Wertung der themenunabhängigen und subjektiven Meinungen in Texten (Esuli und Sebastiani 2006). Methoden und Ansätze dieser Forschungsrichtung ergänzen die thematische Identifizierung der Funktionen durch eine neue Dimension, im Spezifischen durch die Einteilungen der untersuchten Texte in positiv oder negativ.

Die Integration von Methoden aus der NLP erhöht das Potential von nutzergenerierten Daten für die Untersuchung des globalen Städtetzes weiter. Diese weiterführenden Methoden und Ansätzen können in den präsentierten visuell-analytischen methodischen Rahmen integriert werden.

8.2 Methodischer Ansatz

Visuell-analytische Ansätze sind im untersuchten stadtgeographischen Forschungsbereich wenig verbreitet. Die meisten Studien stützen sich auf statistische (Beaverstock et al. 2000b; Taylor et al. 2002a; Derudder et al. 2003) oder mathematisch-netzwerkanalytische Methoden (Alderson und Beckfield 2004; Wall und Van der Knaap 2009). Thierstein und Förster (2008) sowie Rozenblat (Rozenblat 2010), in Zusammenarbeit mit dem „*Laboratoire Bordelais de Recherche en Informatique*“ (LaBRI), haben die aus der kognitiven Forschung bekannte, Vermögen graphischer Darstellungen (Tversky 2000) erkannt. Mit Tulip (LaBRI 2010) hat das LaBRI zudem ein Tool für die Netzwerkanalyse und –visualisierung entwickelt. Wie in Kapitel 2 erläutert fehlt jedoch ein fundierter, methodischer Rahmen für die Untersuchung des globalen Netzwerkes, welcher netzwerk- und die visuell-analytische Methoden integriert. Gestützt auf das *spatialization framework* von Skupin und Fabrikant (2007) wird in dieser Arbeit ein theoriegeleiteter Entwurf eines methodischen Rahmens für die Netzwerkanalyse und –visualisierung des globalen Städtetzes vorgestellt und eine entsprechende Methodik für dessen systematische Erforschung entwickelt.

Das *spatialization framework* von Skupin und Fabrikant (2007) ermöglicht die formale und systematische Integration von mehreren methodischen Ansätzen aus unterschiedlichen Disziplinen. Aufgrund der semantischen Natur der Datengrundlage stützt sich die Datenvorverarbeitung auf computerbasierten semantischen Analysen. Die semantische Generalisierung greift die relationale Natur der stadtgeographischen Prozesse auf und setzt sie unter Berücksichtigung der kognitiven Erkenntnisse (Lakoff und Johnson 1980; Lakoff 1987; Golledge 1995) mittels räumlicher Metaphern um. Die mathematische Transformation stützt sich in Anlehnung an Smith und Timberlake (1995a) auf netzwerkanalytische Methoden. Diese werden durch Ansätze der kartographischen Netzwerkvisualisierung ergänzt. Die geometrische Generalisierung integriert die erprobten Methoden der kartographischen Generalisierung (McMaster 1989; Buttenfield und McMaster 1991) und der Semiotik (Bertin 1974; MacEachren 1995).

Die Ansätze und Methoden aus den unterschiedlichen Disziplinen werden zu einem kohärenten und fundierten Rahmen für die visuell-analytische Erforschung von relationalen Phänomenen integriert. Dabei werden kognitive, netzwerkanalytische und kartographische Konzepte aufeinander abgestimmt und zu einer formalen Grammatik von

graphischen Netzwerkvisualisierungen verflochten. Dies konsolidiert die erste wichtige Eigenschaft des vorliegenden methodischen Rahmens: die Interdisziplinarität.

Die Validität der integrierten Ansätze wird einerseits durch das theoriegeleitete Vorgehen, welches in der Informationsvisualisierung oft fehlt (Couclelis 1998; Skupin 2000), und andererseits dank der systematischen Evaluation der wichtigsten Arbeitsschritte, sichergestellt. Auf die kognitiven und kartographischen Grundsätze wird in dieser Arbeit explizit Bezug genommen. Die kognitiven Aspekte der Visualisierung werden jedoch in der vorliegenden Arbeit nicht evaluiert, sondern es werden die Erkenntnisse aus den empirischen Ansätzen von Fabrikant et al. (2004; 2006) und Montello et al. (2003) umgesetzt. Evaluiert wird insbesondere die Datenfilterung im Zusammenhang mit dem DrL/VxOrd-Algorithmus (Shawn et al. 2011). Dabei werden nicht nur die Skalenabhängigkeit und die Genauigkeit der erzeugten graphischen Darstellungen hervorgehoben, sondern ein systematischer und skalenabhängiger methodischer Ansatz für die Evaluation von Clusterlösungen, welche geographisch inspiriert sind, entwickelt. Ein weiterer wichtiger methodischer Aspekt, der untersucht wird, ist die Klassierung der Akteur-Artikel zu Funktionen. Die inhaltliche Robustheit der mittels des Kategoriensystems in Wikipedia eruierten Funktionen wird mit der Klassierung anhand des *topic modeling* Ansatzes und dem folgenden Vergleich der Resultate evaluiert. Die inhaltlichen Schwerpunkte der Funktionen erweisen sich dabei als methodenunabhängig und robust. Zudem wird der gewählte Modellierungsansatz des Städtensetzes sowohl hinsichtlich der Anzahl betrachteter Städte, als auch der Berechnung der Interaktionsstärke evaluiert. In beiderlei Hinsicht ist die Modellierung robust.

Zusammenfassend präsentiert die vorliegende Arbeit einen interdisziplinären, theoriegeleiteten methodischen Rahmen für die systematische, formale und visuell-analytische Analyse von Netzwerkdaten und ergänzt somit die aktuellen methodischen Ansätze in der stadtgeographischen Forschung. Der methodische Rahmen ist zudem auf weitere Forschungsbereiche, wie die Wissensvisualisierung oder die bibliometrischen Untersuchungen anwendbar. So visualisieren und analysieren beispielsweise Fabrikant und Salvini (2011) die Konferenzinhalte der 25. Internationalen Kartographie Konferenz.

8.2.1 Methodik

Die präsentierte Methodik dient der Untersuchung von grossen, globalen, relationalen und multifunktionalen Städtensetzen, kann aber auch auf ähnlich strukturierte Probleme angewendet werden, wie z.B. die Erkundung von sozialen Medien. Im nächsten Unterkapitel werden die visuell-analytischen Aspekte der Methodik in Bezug zum Forschungsstand am Beispiel der Punkt-Linien Visualisierungen diskutiert. Abschliessend werden am Beispiel der Rollendefinitionen nach Guimera und Almaral (2005a) die spezifischen Herausforderungen der Netzwerkanalyse von geographischen Daten diskutiert.

Visuell-analytische Aspekte

Die Darstellung eines Netzwerkes stellt wegen der hohen Informationsdichte eine besondere Herausforderung dar. Es handelt sich um eine typische kartographische Ausgleichsentscheidung zwischen Informationsdichte und Lesbarkeit. Die kartographische Generalisierung liefert dafür etablierte Ansätze und Methoden, welche auch in dieser Arbeit in das *spatialization framework* integriert werden.

In den Punkt-Linien Visualisierungen wie sie der DrL/VxOrd-Algorithmus generiert (z.B. Abbildung 33) werden Städte mittels zwei Operationen der kartographischen Generalisierung, Aggregation und Zusammenfallen, zu Polen aggregiert. Die Kanten werden zweimal generalisiert: erstens werden die 15 stärksten Beziehungen pro Stadt vor der Dimensionsreduktion selektiert (Klavans und Boyack 2006a), zweitens wird nach der Verräumlichung die Anzahl Kanten mit Hilfe des MST reduziert. Bei den Visualisierungen mittels des GEM-Algorithmus (z.B. Abbildung 39) wird eine reduzierte Zahl an Städten (nur Hubs) und Beziehungen (MST) dargestellt. An dieser Stelle werden insbesondere die Kantenselektion vor der Dimensionsreduktion mit dem DrL/VxOrd-Algorithmus und die Kantenreduktion mittels des MST in beiden Verräumlichungsansätzen diskutiert.

Die Methodenevaluation aus Kapitel 6.2 zeigt, dass die Kantenselektion vor der Dimensionsreduktion mit dem DrL/VxOrd-Algorithmus sowohl die Genauigkeit als auch die räumliche Granularität der Darstellung massgeblich beeinflusst. Die Verräumlichung mit den 15 stärksten Beziehungen pro Stadt weist eine hohe räumliche Granularität auf, sowohl die Makro-Strukturen als auch die Mikro-Strukturen der Daten werden hervorgehoben. Zudem ist die Abbildungsgenauigkeit auf allen Granularitätsebenen hoch (Tabelle 20). Daraus lässt sich schliessen, dass auch die Strukturen des globalen Städtetetzes (z.B. Abbildung 33) adäquat visuell wiedergegeben werden.

Offener ist die Debatte in Bezug auf die Visualisierung des MST. Am Beispiel der Visualisierung des Netzwerkes zwischen den 95 Hubs zeigt sich der Einfluss der Generalisierung: nur 94 der total 4'465 Kanten werden dargestellt. Dies entspricht 2.1% der Beziehungsverhältnisse zwischen den Städten. Nach Chen (1999) ist das *pathfinder network scaling*, welches für die Generierung des MST verwendet wird, eine Methode mit der die wichtigsten Muster und Strukturen im Netzwerk erhalten und hervorgehoben werden. In dieser Hinsicht genügt bereits die Darstellung von 2.1% der Beziehungsverhältnisse für die graphische Vermittlung der entscheidenden Information. Nach Fabrikant et al. (2004) beeinflusst die Darstellung der Trajektorien dennoch die graphische Wahrnehmung von Netzwerkvisualisierungen, die der Distanz-Ähnlichkeitsmetapher folgen. Wenn Trajektorien dargestellt werden, tendieren die Lesenden die Ähnlichkeit zwischen den Knoten aufgrund der Distanz entlang der Pfade und nicht aufgrund der direkten Distanz zu schätzen. In dieser Hinsicht gewinnen die Beziehungsverhältnisse zwischen den Städten, welche mittels einer expliziten Linie dargestellt werden, auf inhaltlicher Ebene enorm an Bedeutung. Konkret wird die hierarchische Struktur des Netzwerkes besonders hervorgehoben. Der MST hebt die Beziehung einer Stadt zu hierarchisch

wichtigeren Städten hervor und ignoriert die Beziehungen zwischen hierarchisch gleichen Städten. Somit veranschaulicht der MST die Strukturen welche mit dem Log-Log-Diagramm (Abbildung 51) eruiert werden. Im Spezifischen kristallisieren sich in beiden Darstellungen die Hauptzentren New York, London und Paris heraus. Der MST, als Resultat des *pathfinder network scalings* mit $r = \infty$, betont diese hierarchische Struktur besonders stark. Weitere, ergänzende Strukturen im Netzwerk werden dadurch herausgefiltert. Eine Berechnung des *pathfinder networks* mit einem kleineren r -Wert, reduziert das Netzwerk weniger stark und weitere sekundäre Strukturen kommen zum Vorschein. In der vorliegenden Arbeit wird das *pathfinder networks* nur mit $r = \infty$ berechnet. Der starken graphischen Hervorhebung der hierarchischen Strukturen im MST wird jedoch die Darstellung der regionalen Zugehörigkeit der Städte mittels Farbton gegenübergestellt (Abbildung 39). Visuell wird damit die Zugehörigkeit der einzelnen Städte zu einem übergeordneten Ganzen hervorgehoben (Fabrikant et al. 2006). Die regionale Zugehörigkeit der Städte ist ein Ausdruck der horizontalen Interaktionen zwischen den Städten und somit die komplementäre Struktur zu den hierarchischen Beziehungen im MST. Detailliertere Analysen über die stärksten Beziehungen in jeder relationalen Region würden die Erkenntnisse über die strukturellen Eigenheiten der einzelnen Regionen und die komplexe Einbettung jeder Stadt im globalen Städtenetz erweitern.

Ein exemplarischer Vergleich einer typischen Netzwerkdarstellung der globalen Städte in dieser Arbeit mit drei typischen Visualisierungen aus der stadtgeographischen Forschung zeigt den konkreten Einfluss des präsentierten visuell-analytischen Ansatzes auf den Informationsgehalt der Darstellungen. Tabelle 26 fasst den Vergleich graphisch zusammen. Alle vier Visualisierungen stellen das globale Städtenetz oder einen Auszug daraus dar. Der erste grosse Unterschied betrifft die explizite graphische Darstellung der interstädtischen Beziehungen mit Trajektorien. Taylor et al. (2002b) (Abbildung 4) verzichten trotz der starken kognitiven Wirkung auf die Wahrnehmung und die inhaltliche Wichtigkeit der Beziehungen im globalen Städtenetz auf die Darstellung derselben. Demgegenüber steht die Visualisierung von Alderson und Beckfield (2004) (Abbildung 8), die einer klassischen sozialen Netzwerkdarstellung entspricht. Da die Beziehung ohne Kantenreduktion und Generalisierung dargestellt werden, sind sie nicht voneinander unterscheidbar und nicht lesbar. Sowohl die schematische Darstellung nach Rozenblat (2004) (Abbildung 9) als auch die Punkt-Linien Visualisierung dieser Arbeit (Abbildung 30) stellen eine reduzierte und lesbare Anzahl Beziehungen dar und heben zudem die Beziehungsstärke hervor. Punkt-Linien Visualisierungen, in denen die graphische Darstellung interaktiv generalisiert werden kann, würden den vorgelegten methodischen Rahmen weiter ergänzen. Mit empirischen Probandenstudien könnte zudem die Wahrnehmung der Punkt-Linien Visualisierungen empirisch überprüft werden.

Tabelle 26: Vergleich des Informationsgehalts verschiedener Darstellungen des globalen Stdtenetzes

Semantische Primitive	Graphische Variable	Punkt-Linien Visualisierung (Abbildung 33)	Schematische Darstellung nach Rozenblat (2004) (Abbildung 9b)	Netzwerkvisualisierung nach Alderson und Beckfield (2004) (Abbildung 8)	Kartogramm nach Taylor et al. (2002b) (Abbildung 4)
Lokus		Stadt	Stadt	Stadt	Stadt
Lage		Stellung im Netzwerk	Stellung im Netzwerk	Stellung im Netzwerk	Geographische Lage
Grösse		Zentralitt	-	-	-
Farbe (Farbton oder Helligkeit)		Regionale Zugehrigkeit	Zentralitt	Regionale Zugehrigkeit	Zentralitt
Form		Rolle	-	-	-
Beschriftung		Stadtname	Stadtname	-	Stadtname
Trajektorie		Interstdtische Beziehung	Interstdtische Beziehung	Interstdtische Beziehung	-
Grösse		Beziehungsstrke	Beziehungsstrke	-	-
Farbe (Helligkeit)		Beziehungsstrke	-	-	-

Die Lage der Stdte in der Visualisierung nach Taylor et al. (2002b) Abbildung 4 entspricht als einzige der geographischen Lage. In den anderen drei Visualisierungen hebt die Lage die topologische Stellung im Netzwerk hervor und unterstreicht somit diese zentrale Eigenschaft der Stdte aus relationaler Perspektive. Im Gegensatz zu der in dieser Arbeit entwickelten Visualisierungsform wird bei den drei anderen Visualisierungen, die den aktuellen Forschungsstand reprsentieren, nur die Farbe als Farb-Muster-Variable verwendet. Damit wird das Potential einer effektiven und konsequenten Anwendung der graphischen Variablen nicht ausgeschpft und der Informationsgehalt bleibt bescheiden.

Neben Netzwerkvisualisierungen werden in der vorliegenden Arbeit auch SOM zur visuell-analytischen Untersuchung des globalen Stdtenetzes eingesetzt. SOM sind in der Erforschung des globalen Stdtenetzes eine Neuigkeit und werden in dieser Arbeit insbesondere fr die Analyse der Funktionen eingesetzt. SOM haben zwei entscheidende Vorteile: einerseits ist es mglich, grosse Mengen an multidimensionalen Daten zu organisieren und zu visualisieren (Brner et al. 2003) und andererseits erfllen SOM die Container-Metapher optimal. Dieser letzte Punkt zeigt sich insbesondere in Abbildung 45: die Zugehrigkeit einer Stadt zu einer Funktion wird graphisch mit dem Zwei-Ebenen-System hervorgehoben.

Die vorliegende Arbeit leistet zudem eine Weiterentwicklung der Methodik der SOM. Die herkömmlichen SOM haben ein starres, regelmässiges Raster mit hexagonalen Neuronen und folgen der Distanz-Ähnlichkeitsmetapher damit nur beschränkt. Um diesen visuellen und kognitiven Nachteil teilweise zu kompensieren werden Neuronen mittels Kartogramm aufgrund der U-Matrix, welche die Ähnlichkeit der benachbarten Neuronen wiedergibt, verformt. Somit werden die Ähnlichkeit und die graphische Distanz in der SOM kongruent dargestellt; ähnliche Neuronen rücken näher zueinander und unähnliche Neuronen distanzieren sich voneinander. Die latenten Strukturen der Daten werden dadurch salienter hervorgehoben.

Zusammenfassend zeigt die vorliegende Arbeit, dass die theoriegeleitete und kartographisch inspirierte Visualisierung von Netzwerkdaten die Lesbarkeit und den Informationsgehalt gegenüber den gängigen Netzwerkdarstellungen aus dem betrachteten stadtgeographischen Forschungsbereich massgeblich erhöht. Der wichtigste Vorteile von visuell-analytischen Ansätzen ist der erleichterte Zugang zu grossen Datenmengen auch für nicht-technische Nutzer und die damit verbundene Erkennung von Mustern und Strukturen in den Daten (Börner et al. 2003). Zwei Beispiele aus der vorliegenden Arbeit illustrieren diesen Vorteil. Die graphische Darstellung des globalen Städtetzes (z.B. Abbildung 39) ermöglicht die gleichzeitige Betrachtung von Detailinformationen betreffend einer Stadt (z.B. hohe Zentralität von Paris und Zugehörigkeit zur WESTEUROPÄISCHEN Region) und deren Kontext (z.B. gleichhohe Zentralität wie London und New York und überdurchschnittliche Zentralität verglichen mit den weiteren WESTEUROPÄISCHEN Städten). Diese doppelte Betrachtungsebene erleichtert die Erkennung von Mustern und die Interpretation der Daten. Kapitel 6.1 illustriert den Vorteil von Visualisierungen beim Vergleichen von Resultaten. Die Korrelationsanalysen zeigen die Übereinstimmung zwischen dem Referenznetzwerk und dem englischsprachigen Städtetz deutlich auf. Jedoch nur durch die graphische Darstellung der Daten in der Form eines Netzwerkes (z.B. Abbildung 20) oder eines Scatterplots (z.B. Abbildung 22) lassen sich die Gemeinsamkeiten und Unterschiede zwischen den Resultaten vollständig eruieren. In Abbildung 20 ist die unterschiedliche Vernetzung der OST-/SÜDOSTASIATISCHEN Städte in beiden Netzwerken erkennbar. Abbildung 22 hebt die sprachlich-räumlichen Verzerrungen in den Daten hervor, die aufgrund der Korrelationsanalysen nicht erkennbar sind.

Definition der Rollen im globalen Städtetz

Nach Guimera und Almaral (2005a) ist die Definition der Rollen universell und eignet sich somit für alle Arten von Netzwerken, seien sie aus dem biologischen, physischen oder sozialen Bereich. Die Anwendung auf das Städtetz mit dem Ziel, relational wichtige Städte von den anderen zu unterscheiden, hat sich als nützlich erwiesen. Da bei der Rollendefinition die Zentralität einzelner Städte mit der Zentralität von Städten aus der gleichen Region verglichen wird und somit die Definition auf normalisierten Werten beruht, ermöglicht diese Methode das sprachbedingte Übergewicht der US-amerikanischen Städte auszugleichen.

Wegen der räumlichen Eigenschaften der Daten sind bei der Anwendung der Rollendefinition auf das globale Städtetz auch geographische Konzepte eng damit gekoppelt. Da die Rollendefinition von den identifizierten Regionen abhängt, hat diese mit den aus der Analyse von räumlichen Daten bekannten Herausforderungen umzugehen. Insbesondere die Abgrenzung und Grösse der Region bestimmen die Definition der Hubs in dieser Region. Ein illustrierendes Beispiel dafür sind die vier Hubs in der PHILIPPINISCHEN Region (Baguio, Cebu, Quezon, Davao). Diese Städte mit einer sehr niedrigen Zentralität sind nur Hubs, weil die PHILIPPINEN eine kleine und eigenständige Region bilden. Wenn die Region Teil der grossen OST-/SÜDOSTASIATISCHEN Region wäre, wäre Manila der einzige Hub auf den Philippinen. In Europa werden auch drei kleinere Regionen ausgeschieden: RUMÄNIEN, GROSSBRITANNIEN/IRLAND und die TÜRKEI. In diesen drei Regionen ist die methodisch-bedingte Verzerrung geringer. Die hervorgehobenen Hubs, Budapest, London, Ankara und Istanbul, würden auch in einer grossen umfassenden europäischen Region als Hubs gelten. Abgesehen von der Situation der philippinischen Städte ist die Taxonomie für die vorliegende Arbeit valide.

Neben der Differenzierung der Städte in Hubs und Nicht-Hubs, lassen sich sowohl Hubs als auch Nicht-Hubs basierend auf dem Partizipationskoeffizient weiter unterscheiden. Die Unterscheidung der Nicht-Hub-Städte in ultra-periphere (R1), periphere (R2), verwandtschaftslose Knoten (R4) und Verbinder (R3) wird in dieser Arbeit nicht weiter berücksichtigt. Diese Unterscheidung ist jedoch nach Bruggmann (2012) auch stark von der geographischen Lage der Städte abhängig. Die Rolle als verwandtschaftsloser Knoten (R4) ist öfters der Ausdruck der grenznahen geographischen Lage der Städte. Zwei Beispiele dafür sind Abu Dhabi in der OST-/SÜDOSTASIATISCHEN Region und Tripolis in der WESTEUROPÄISCHEN Region. Beide Städte befinden sich an der geographischen Grenze der eigenen Region und in einem geographischen Gebiet, in dem viele relationale Regionen zusammenkommen. Im Spezifischen umfasst dieses geographische Gebiet die AFRIKANISCH/NAHÖSTLICHE, die TÜRKISCHE, sowie die WESTEUROPÄISCHE und die OST-/SÜDOSTASIATISCHE Region. Somit ist der Partizipationskoeffizient dieser Städte hoch, obwohl sie vor allem Beziehungen zu geographisch benachbarten Städten haben. Städte, die sich geographisch im Zentrum einer Region befinden und ein Beziehungsmuster mit einer zu Abu Dhabi oder Tripolis äquivalenten geographischen Reichweite haben, zählen wegen des geringen Partizipationskoeffizienten zu den peripheren Knoten.

Ein zweites wichtiges geographisches Konzept sind Skalenebenen. Sowohl die Hubdefinition mit den Rollen nach Guimera und Almaral (2005a) als auch die Regionalisierung anhand des Blondel-Algorithmus tragen dieser geographisch inhärenten, mehr-skaligen Eigenschaft Rechnung. Sowohl die Hubdefinition als auch die Identifizierung der Regionen beziehen sich in dieser Arbeit jedoch ausschliesslich auf die höchste Aggregationsstufe. Die stadtgeographischen Phänomene werden somit ausschliesslich auf höchster geographischer Skala untersucht, obwohl vom Thema her eine umfassende Untersuchung gleichzeitig mehrere Ebenen berücksichtigen sollte. Das Potential einer mehr-skaligen Analyse des Städtetetzes mit den gewählten Methoden kann jedoch am

Beispiel der Region OZEANIEN/SÜDAFRIKA gezeigt werden. Auf höchstem Skalenniveau wird diese Region über die zwei australischen Hubs, Sydney und Melbourne, mit dem Rest des Netzwerks verbunden. Auf tieferer Skala teilt sich diese Region in eine OZEANISCHE und eine SÜDAFRIKANISCHE Region auf. In Anbetracht der veränderten Skala des untersuchten stadtgeographischen Phänomens kommt neben Sydney und Melbourne auch Johannesburg als Hub hervor.

8.3 Globales Städtenetz

Die Diskussion des thematischen Aspekts der Arbeit tangiert drei Punkte. Erstens werden die regionale Struktur und die daraus entstehenden Grenzen im Netzwerk diskutiert. Über diesen Aspekt liegen noch wenige Erkenntnisse, insbesondere auf tiefer Massstabsebene vor. Der zweite Punkt behandelt die Stellung der einzelnen globalen Städte in Bezug zum Forschungsstand. Im letzten Teil werden die unterschiedlichen Funktionen und funktionalen Netze diskutiert. Die Resultate zum multifunktionalen globalen Städtenetz sind in dieser Vollständigkeit und mit diesem Detaillierungsgrad ein Novum und füllen eine wichtige inhaltliche Lücke.

Relationale Regionen und Grenzen

Die wohl wichtigste Erkenntnis über die Organisation des globalen Städtenetzes aus geographischer Perspektive ist die wichtige Rolle, die der Raum auch im Zeitalter der Globalisierung immer noch spielt. Zwar sind nicht die Beziehungen zwischen den räumlich nahe liegenden Städten die stärksten, trotzdem haben die relationalen Regionen bzw. die einzelnen Städtesysteme, eine starke, räumlich differenzierte Prägung. Diese entscheidende Charakteristik der relationalen Geographie wird insbesondere durch die Projektion der relationalen Regionen auf eine Weltkarte (Abbildung 26) deutlich. Die relationalen Regionen decken sich mit den Kontinenten und Subkontinenten. Auf einem tieferen, in dieser Arbeit nicht weiter untersuchten Aggregationsniveau, ist die starke räumliche Prägung gleichwohl vorhanden. Dies entspricht den empirischen Ergebnissen von Taylor et al. (2002a), Derudder et al. (2003) und Rozenblat et al. (2010), die mit anderen Daten und anderen Klassierungsmethoden zu ähnlichen Resultaten kommen.

Die Resultate zeigen, dass unabhängig davon, ob das betrachtete Städtenetz allgemein oder funktional differenziert ist, mit allen Städten oder nur mit den Hubs modelliert wird, mehrheitlich stabile relationale Regionen identifiziert werden. Mit wenigen Ausnahmen kristallisieren sich WESTEUROPA, KANADA, USA, OST-/SÜDOSTASIEN und OZEANIEN/SÜDAFRIKA als zentrale sowie SÜD-/ZENTRALAMERIKA, SÜDASIEN, und zum Teil AFRIKA/NAHER OSTEN, als periphere Regionen heraus. Dies entspricht den wenigen bisherigen Erkenntnissen über die relationale regionale Organisation (Rozenblat et al. 2010). Die vorliegenden Resultate lassen jedoch eine differenziertere Aussage zu und ermöglichen somit eine detailliertere Diskussion, die neue Erkenntnisse insbesondere über die inneren und äusseren Grenzen der Regionen bringt. Der amerikanische Kontinent ist mit NORDAMERIKA auf der einen Seite sowie ZENTRAL-/SÜDAMERIKA auf der anderen Seite, durch eine herausragende Stabilität der regionalen

Organisation geprägt. Während NORDAMERIKA sich teilweise in die USA und KANADA unterteilt, aber immer zentral und stark integriert ist, sticht SÜD-/ZENTRALAMERIKA aufgrund der peripheren Stellung und der inneren Fragmentierung hervor. Die erste Trennlinie verläuft zwischen den Nordamerika-orientierten Städten in Zentralamerika und den Europa-orientierten südamerikanischen Städten. Eine zweite prägende Grenze verläuft zwischen Brasilien und den anderen südamerikanischen Städten.

Obwohl Europa auch in relationaler Perspektive eine zentrale Stellung innehat und die Städte durchaus stark integriert sind, lassen sich innere Grenzen erkennen. Wie Derudder et al. (2003) erkennen, erstreckt sich eine Grenze zwischen Kontinentaleuropa und den Britischen Inseln sowie eine andere zwischen West- und Osteuropa. Derudder et al. (2003) sprechen zudem von einer deutsch-französischen Grenze, die in den vorliegenden Resultaten aber nicht erkennbar ist. Hingegen grenzen sich je nach Analyse insbesondere Russland und die europäischen Anrainerstaaten, sowie die Türkei und Rumänien ab. Seltener bildet die iberische Halbinsel eine gemeinsame Region mit ZENTRAL-/SÜDAMERIKA. Labiler sind die südliche und die östliche Aussengrenze der EUROPÄISCHEN Region. Östlich schliesst die innerste Grenze Rumänien, Ukraine, Weissrussland, Russland und die Türkei aus der europäischen Region aus. Russland, die Ukraine und Weissrussland bilden eine gemeinsame Region mit den zentralasiatischen Städten. Rumänien und die Türkei sind eigenständig. St. Petersburg, Moskau, Ankara, Istanbul und Bukarest sind als einzige Hubs dieser Regionen in die EUROPÄISCHE Region integriert. Nach Süden schliesst die innere EUROPÄISCHE Grenze Libyen und Tunesien mit ein. Die äusserste Grenze schliesst ganz Afrika und den nahen Osten in die EUROPÄISCHE Region ein. Die Auflösung der AFRIKANISCH/NAHÖSTLICHEN Region und deren Integration in die EUROPÄISCHE Region entspricht den Resultaten von Rozenblat et al. (2010). Die relationale Geographie Afrikas ist jedoch komplex und lässt sich nicht mit der simplen Integration in Europa abhandeln. Ein wichtiges Element der Grenze zwischen Europa und Afrika ist, wie oft in der Geschichte, das Mittelmeer, welches nicht nur als Grenze sondern oft als Bindeglied zwischen den Anrainerstaaten wirkt. Wie oben erwähnt bindet das Mittelmeer die tunesischen und libyschen Städte und teilweise den ganzen nordafrikanischen Raum an Europa.

Obwohl verschiedene afrikanische Städte zu einer Region gehören, gibt es den afrikanischen Kontinent in der relationalen Geographie nicht. Abgesehen von der unscharfen Grenze zu Europa gehört im Süden auch Südafrika nicht zur AFRIKANISCH/NAHÖSTLICHEN Region. Die südafrikanischen Städte bilden eine gemeinsame Region mit Ozeanien. Im Osten weitet sich hingegen die AFRIKANISCH/NAHÖSTLICHE Region meistens in den nahen Osten aus und positioniert sich, stark fragmentiert, immer peripher zu Europa. Zwei wichtige, in sich wiederum fragmentierte und unscharf abgegrenzte Subregionen lassen sich dabei unterscheiden: der arabische Raum (Nordafrika und naher Osten) und die Sub-Sahara.

SÜDASIEN, OST-/SÜDOSTASIEN und die PHILIPPINEN sind drei eigenständige Regionen. Aus regionaler Perspektive sind die PHILIPPINISCHEN Städte klar von den anderen OST-

/SÜDOSTASIATISCHEN Städten abgekoppelt und auch untereinander schwach integriert. Aus städtischer Perspektive sind die vier PHILIPPINISCHEN Hubs Teil der OST-/SÜDOSTASIATISCHEN Region. Die OST-/SÜDOSTASIATISCHE Region unterteilt sich zudem in eine japanisch-koreanische und eine chinesisch-südostasiatische Region. OST-/SÜDOSTASIEN nimmt, wie auch von der GaWC-Gruppe (Taylor et al. 2002a; Derudder et al. 2003) eruiert, eine zentrale Stellung ein. SÜDASIEN und die PHILIPPINEN sind hingegen peripherer und in sich fragmentiert.

Die ozeanischen und die südafrikanischen Städte bilden eine gemeinsame Region und bilden somit einen Teil der Commonwealth-Region nach Derudder et al. (2003). Entscheidender Unterschied sind aber die kanadischen Städte: nach Derudder et al. (2003) gehören sie auch zur Commonwealth-Region. Nach den vorliegenden Resultaten ist die Integration der kanadischen Städte in dieser OZEANISCH/SÜDAFRIKANISCHEN Region gering.

Die relationalen Regionen sind nur ein Organisationselement im globalen Städtenetz. Das folgende Unterkapitel betrachtet darüber hinaus die einzelnen globalen Städte und deren hierarchische Organisation.

Globalisierte und globale Städte

Die Unterscheidung zwischen globalen Städten und der Gesamtheit der globalisierten Städte ist in Anlehnung an Marcuse und van Kempen (2000) nicht prioritär, da aus dieser Perspektive die globalen Städte einfach den Kern des globalen Städtenetzes darstellen. Die Abgrenzung der globalen Städte ist unscharf und hat eine willkürliche Dimension. Die Abgrenzung der globalen Städte mit der Hubdefinition nach Guimerà und Amaral (2005a) erzeugt eine scharfe Grenze, wo es eigentlich keine gibt und ist somit teilweise diskutabel. Dieses Unterkapitel erörtert die verwendete Definition der globalen Städte, indem sie mit den Ergebnissen aus dem Stand der Forschung verglichen werden. Zu beachten ist jedoch, dass die globalen Städte in den Resultaten explizit multifunktional sind, die Erkenntnisse basierend auf dem Stand der Forschung sich hingegen mehrheitlich auf ökonomische Daten stützen. Ein Vergleich ist ansatzweise trotzdem möglich. Anhang 9 fasst die Resultate dieser Arbeit und die Erkenntnisse aus dem Stand der Forschung über die Zentralität der 95 Hubs zusammen. Für jede Stadt wird die Rangierung aufgrund der Zentralität im gesamten Städtenetz (zwischen den 3'532 Städten), im Städtenetz zwischen den 95 Hubs und in den acht funktionalen Netzwerken angegeben. Tabelle 27 greift 18 Städte aus Anhang 9 heraus. Entweder unterscheiden sich für diese Städte die Resultate dieser Arbeit besonders vom Forschungsstand oder sie sind exemplarische Beispiele für spezifische Muster. Dabei wird für jede Stadt die Zentralität im gesamten Netzwerk, im Städtenetz zwischen den Hubs und im WIRTSCHAFTLICH/TECHNOLOGISCHEN Netzwerk hervorgehoben. Das wirtschaftlich/technologische Netzwerk stimmt inhaltlich am Besten mit den Städtenetzen aus der aktuellen stadtgeographischen Forschung überein. In den nächsten Abschnitten wird unter anderem auf die Stellung der 18 Städte in Tabelle 27 eingegangen. Zuerst wird jedoch die Anzahl globaler Städte und deren weltweite Verteilung diskutiert.

Tabelle 27: Gegenüberstellung der Resultate der Arbeit und des Forschungsstandes. Resultate: Rangierung der 95 Hubs im allgemeinen und im wirtschaftlich/technologischen Netzwerk. Stand der Forschung: Kategorisierung und Rangierung der 95 Hubs nach den gleichen Quellen und Grundsätzen wie in Kapitel 2.3 (siehe Tabelle 3, Anhang 9 und Anhang 1)

	Eigene Resultate			Global Cities				World Cities					
	Allgemeines Netzwerk		Funktionale Netzwerke	GaWC-Forschungsgruppe	Wall und v.d. Knaap				Alderson und Beckfield				Friedmann
Stadt	3'532 Städte	95Hubs	Wirtschaft/Technologie	Global Connectivity (2010)	OD (2005)	ID (2005)	OD (2005)	ID (2005)	OD (2000)	ID(2000)	OD (2007)	ID (2007)	1995
Europa													
London	1	1	2	α++	3	2	4	2	4	2	4	2	Gfa
Paris	3	3	3	α+	2	5	2	5	3	3	2	3	Ina
Berlin	7	7	18	β+	27								
Munich	19	20	25	α-	13	26	7	28	8	20	5		Ra
Rome	8	8	24	β+					30				
Milan	27	26	19	α		7		9		14			Ra
Madrid	22	21	23	α		13		8	45	12		15	Ina
Barcelona	29	27	29	α-				21		40			Mna
Vienna	12	12	26	α-				17		35			
Moscow	10	9	15	α									
St. Petersburg	34	28	32	γ									
Nordamerika													
New York	2	2	1	α++	1	1	1	1	2	1	1	1	Gfa
Washington	6	6	8	α						23			
Ost- und Südostasien													
Hong Kong	21	13	4	α+		4		4		10		8	Ra
Tokyo	15	10	11	α+	8	6	3	6	1	4	3	4	Gfa
Beijing	17	15	14	α					24	36			
Singapore	33	19	10	α+	30	3		3		9		11	Mna
Shanghai	53	33	12	α+									
Philippinen													
Manila	85	45	33	β+					42				
Quezon	331	91	91										
Davao	444	93	94										
Cebu	490	94	93										
Baguio	567	95	95										

Von den 3'532 betrachteten Städten sind 95 globale Städte. Friedmann (1986, 1995) spricht von 30 globalen Städten, Sassen (1991) von drei globalen Städten und die

GaWC-Forschungsgruppe (GaWC 2011) identifiziert für 2010 um die 200 globale Städte. 95 globale Städte entsprechen somit ungefähr der erwarteten Grössenordnung. Zudem wird damit die hierarchische Struktur des globalen Städtenetzes nach Friedmann (1986) und Sassen (1991) bestätigt. In den vorliegenden Resultaten entsprechen die globalen Städte 3% der betrachteten Städte, die aber eine überdurchschnittliche Zentralität auf sich vereinen: 13% der nähebasierten Zentralität, 32% des gewichteten *degrees* und gar 54% der *betweenness* Zentralität. Dies unterstreicht wiederum die hierarchische Struktur.

Die geographische Verteilung der globalen Städte hebt neue und veränderte Geographien hervor. Von den 95 Städten befinden sich 35% in Asien, 26% in Europa, 18% in Süd- und Zentralamerika, 9% in Nordamerika, 6% in Afrika und 2% in Ozeanien. Damit unterscheidet sich die geographische Verteilung markant von den vorgängigen Studien von Alderson und Beckfield (2004), Friedmann (1986, 1995) sowie der GaWC-Forschungsgruppe (2009b, 2009a, 2010, 2011) und van der Knapp (2011). Europa als traditioneller Kontinent mit dem höchsten Anteil globaler Städte wird in der vorliegenden Arbeit von Asien (35% der globalen Städte) überholt. Zudem ist der höhere Anteil (18%) an süd- und zentralamerikanischen und der geringere Anteil (9%) an nordamerikanischen Städten bemerkenswert. Mit einem Anteil von 6% sind die afrikanischen Städte im Vergleich zum Stand der Forschung übervertreten. Der Grund für diese veränderte Verteilung liegt viel mehr in der gewählten Definition, als in realen Veränderungen. Sowohl die GaWC-Forschungsgruppe (2009b, 2009a, 2010, 2011) als auch Alderson und Beckfield (2004) definieren die globalen Städte einzig aufgrund der Zentralitätswerte. Bei der Betrachtung des gewichteten *degrees* und der nähe-basierten Zentralität entspricht die geographische Verteilung dem bisherigen Stand der Forschung: sehr hohe Anteile an globalen Städten in Nordamerika (ca. 35%) und Europa (ca. 40%), mittlere Anteile in Asien (ca. 15%) und geringe in Südamerika (ca. 5%), Ozeanien (ca. 3%) und sehr geringe in Afrika (1 %). Die geographische Verteilung aufgrund der *betweenness* Zentralität entspricht in der Tendenz der Verteilung der Hubs und hebt insbesondere die hohe *betweenness* Zentralität der asiatischen Städte und die geringe *betweenness* Zentralität der nordamerikanischen Städte hervor.

Das in dieser Arbeit eingesetzte Konzept der Hubs bringt einen erheblichen Mehrwert verglichen mit der simplen Charakterisierung der globalen Städte mittels eines (GaWC 2009b, 2009a, 2010, 2011) oder mehrere (Alderson und Beckfield 2004) Zentralitätsmasse. Die Hubdefinition vereinigt in sich sowohl die regionale Organisation, als auch die lokale und globale Integration der Städte. Mit dieser Definition wird der Komplexität und insbesondere der mehrskaligen Rolle der globalen Städte zwischen lokaler und globaler Integration besser Rechnung getragen. Städte mit ausschliesslich hoher lokaler Integration, z.B. die NORDAMERIKANISCHEN Städte, verlieren mit der Hubdefinition an Bedeutung. Städte die auf mehreren Skalen integriert sind, z. B. die OST/SÜDOSTASIATISCHEN Städte, gewinnen an Bedeutung. Die Ergänzung der Hubdefinition mit den Zentralitätswerten (siehe Abbildung 39 und Anhang 2) bringt weitere Muster zur Erscheinung, z.B. dass die NORDAMERIKANISCHEN globalen Städte zwar in

ihrer Anzahl untervertreten sind, jedoch aufgrund ihrer hohen Zentralität eine erhebliche Bedeutung im Netzwerk haben. Für SÜD-/ZENTRALAMERIKA gilt genau das Gegenteil. Die von Sassen (1991) postulierte neue Geographie, die das Zentrum-Peripherie-Gefälle sprengt, ist in den vorliegenden Resultaten, wie in anderen empirischen Arbeiten, nicht ersichtlich. NORDAMERIKANISCHE, EUROPÄISCHE, OST-/SÜDOSTASIATISCHE Städte bleiben die dominanten Zentren des globalen Städtensetzes.

Bei der Betrachtung der einzelnen globalen Städte pro Kontinent ergeben sich zweideutige Resultate: einerseits werden die grundsätzlichen Erkenntnisse bestätigt, andererseits rückt die Stellung bestimmter Städte in ein neues Licht.

In Europa zeigen die Städte hinsichtlich ihrer Zentralität eine klare Hierarchie: Paris und London sind an der Spitze, gefolgt von mehreren europäischen Hauptstädten wie z.B. Berlin, Rom, Madrid und Wien. Während die zentrale Stellung von London und Paris grundsätzlich mit dem Stand der Forschung übereinstimmt, unterscheidet sich die Stellung der Hauptstädte vom Stand der Forschung: die Hauptstädte (Rom, Berlin, Madrid) weisen in den vorliegenden Resultaten sowohl im allgemeinen als auch im WIRTSCHAFTLICH/TECHNOLOGISCHEN Netzwerk eine höhere Zentralität als weitere Städte im gleichen Land auf. Im Stand der Forschung ist die Zentralität der Hauptstädte, insbesondere Roms und Berlins, sehr gering. Der Vorsprung der Hauptstädte im allgemeinen Netzwerk ist mit der verwendeten Datengrundlage intuitiv erklärbar: der Einbezug politischer, kultureller und historischer Daten in Wikipedia stärkt die Hauptstädte gegenüber anderen Städten. Die Zentralität der Hauptstädte im WIRTSCHAFTLICH/TECHNOLOGISCHEN Netzwerk gilt hingegen als eine neue Erkenntnis. Nur in Italien rangiert, im Einklang mit dem Forschungsstand, Mailand vor Rom im WIRTSCHAFTLICH/TECHNOLOGISCHEN Netzwerk. Die im Vergleich zum Stand der Forschung höhere Zentralität von Berlin, Moskau und St. Petersburg stimmt mit dem Postulat von Friedmann (1995) betreffend einer zunehmenden Zentralität dieser Städte überein.

Die globalen Städte des deutschsprachigen Raumes, Berlin, München und Wien, unterscheiden sich markant von der diskutierten Literatur, wo München, Zürich, Frankfurt und Düsseldorf erwähnt werden. Im Falle Berlins spielen unterschiedliche Faktoren eine Rolle; einerseits hat sich Berlin in den letzten zehn bis fünfzehn Jahren, also der Zeitspanne zwischen den Ergebnissen von Friedmann (1995), Alderson und Beckfield (2004) und den vorliegenden Resultaten, stark ins europäische Städtensetz integriert, andererseits ist die historische und politische Zentralität von Berlin als Hauptstadt Deutschlands, die in der erwähnten Literatur nicht berücksichtigt wird, unbestritten. Zürich und Frankfurt verdanken ihre Zentralität hingegen insbesondere dem Finanzplatz. In der *global city*-Definition ist eben dieser ökonomische Sektor das entscheidende Kriterium, wohingegen es im multifunktionalen Ansatz dieser Arbeit bloss ein Element unter mehreren ist. Dadurch erklärt sich die Abschwächung von Frankfurt und Zürich in der vorliegenden Arbeit. Die in der vorliegenden Arbeit tiefere Zentralität Düsseldorfs ist vermutlich durch räumliche Faktoren gegeben. Insbesondere von Friedmann (1986, 1995)

wird Düsseldorf als Überbegriff für die ganze Region Rhein-Ruhr verwendet. Diese Region kristallisiert sich nach Thierstein et al. (2011) zu einer *mega-city-region*, welche neben Düsseldorf, auch Bochum, Bonn, Dortmund, Duisburg, Essen und Köln umfasst. Die Analysen in Wikipedia betrachten jede dieser sieben Städte einzeln und nicht als eine integrierte *mega-city-region*. Dadurch ist die Zentralität von Düsseldorf in dieser Arbeit gering.

In Nordamerika ist die Übereinstimmung zwischen den Resultaten und dem Stand der Forschung relativ gross. Insbesondere die weltweite Rangierung New Yorks an erster oder zweiter Stelle ist unbestritten. Die Resultate unterscheiden sich jedoch insbesondere hinsichtlich der Zentralität von Washington. In Analogie zu den europäischen Hauptstädten wird Washington durch die POLITISCHE Funktion gestärkt, was auch zu einer höheren Zentralität im allgemeinen Netzwerk führt. Die Zentralität Miamis in Bezug auf den Forschungsstand ist kontrovers. Die Resultate dieser Arbeit bestätigen die Erkenntnisse von Alderson und Beckfield (2004), dass Miami nicht die Stellung einer globalen Stadt innehat.

In Bezug auf den asiatischen Kontinent unterscheiden sich die Resultate vom Stand der Forschung insbesondere in der grösseren Anzahl der globalen Städte und der grösseren räumlichen Ausdehnung der Region. Vergleichbar ist hingegen die hierarchische Organisation: Tokyo und Hongkong stehen an der Spitze, gefolgt in erster Linie von Peking, Singapur sowie Shanghai und in zweiter Linie von Mumbai, Delhi, Manila, Kuala Lumpur, Bangkok, Seoul und Osaka. Singapur und Shanghai, aber auch Hong Kong und Peking haben eine ausgeprägt höhere Zentralität im WIRTSCHAFTLICH/TECHNOLOGISCHEN Netzwerk. In räumlicher Hinsicht treten in dieser Arbeit die nächstlichen globalen Städte (Bagdad, Beirut, Damaskus, Jerusalem, Tel Aviv und Teheran) und eine grössere Anzahl an indischen Städten (Kalkutta, Chennai, Bangalore, Karachi, Hyderabad, Lahore, Pune) neu hervor. Die Zentralität dieser Städte ist insbesondere von der POLITISCHEN und der SOZIO/HISTORISCH/MILITÄRISCHEN Funktion gegeben. Die philippinischen Städte erfordern eine besondere Diskussion. Neben Manila zählen auch vier weitere kleine Städte, die in der Literatur keine Erwähnung finden, zu den globalen Städten. Ihre globale Zentralität ist jedoch in jeder Hinsicht gering bzw. unbedeutend; im gesamten Städtenetz rangieren sie gar hinter dem Rang 300. Diese Anomalie hat wie vorgängig erwähnt methodische Gründe.

Die Erkenntnisse über Ozeanien sind konform zur betrachteten Literatur. Sydney und Melbourne zählen zu den globalen Städten und haben eine vergleichbare Zentralität zu den mittleren nordamerikanischen oder europäischen globalen Städten.

Die süd- und zentralamerikanischen Städte unterscheiden sich zwar hinsichtlich ihrer Anzahl von den gemäss Forschungsstand ersichtlichen Zahlen, jedoch nicht hinsichtlich ihrer Zentralität. Aus der Gesamtperspektive sind die süd- und zentralamerikanischen globalen Städte peripher. Buenos Aires, Sao Paulo und Mexiko City sind erwartungsgemäss die zentralsten Städte dieser Region. Gefolgt werden sie von Havanna, Rio und Santiago. Havanna und Rio werden in der bisherigen Literatur nur am Rande erwähnt.

Obwohl Karakas und Brasilia Hauptstädte sind und somit eine wichtige politische Funktion haben, ist ihre Zentralität gering.

In Afrika werden einzig und zudem nur in wenigen Beiträgen Pretoria, Johannesburg und Kairo zu den globalen Städten gezählt. Die Resultate weisen sechs afrikanische Städte aus. Ihre Zentralität ist aber sehr gering. Kairo ist im Einklang mit der GaWC-Forschungsgruppe (2011) die zentralste afrikanische Stadt. Alexandria, Dakar, Algier, Lagos und Nairobi sind dagegen peripher. Pretoria und Johannesburg zählen trotz der, im Vergleich zu den anderen afrikanischen Städten, hohen Zentralität nicht zu den globalen Städten. Der Grund dafür ist ihre Integration in die OZEANISCH-SÜDAFRIKANISCHE Region. Verglichen mit Sydney, Melbourne, aber auch mit anderen ozeanischen Städten, sind die zwei südafrikanischen Städte peripher.

Bei einer globalen Betrachtung der Städte, insbesondere was den Kern des globalen Netzwerkes betrifft, fällt die Diskussion unterschiedlich aus. Erstens, stimmen die Resultate in Bezug auf die herausragende Zentralität von London und New York mit dem Stand der Forschung überein: New York ist die WIRTSCHAFTLICH/TECHNOLOGISCH zentralste Stadt, London ist im allgemeinen Netzwerk die zentralste Stadt. Zweitens, stufen insbesondere Sassen (1991), Friedmann (1986) und Alderson und Beckfield (2004; Alderson et al. 2010) Tokyo neben London und New York als dritte zentralste Stadt im Netzwerk ein. Die vorliegenden Resultate bestätigen diese Erkenntnisse nicht, sondern stimmen eher mit Wall und van der Knapp (2011) und der GaWC-Forschungsgruppe (2011) überein und weisen für Tokyo eine Zentralität im Rang zehn auf. Die Zentralität von Tokyo ist somit vergleichbar mit derjenigen von Hong Kong, Peking und teilweise Singapur und Shanghai. Die OST-/SÜDOSTASIATISCHE Region profiliert sich somit als polyzentrische Region, in der sich Tokyo, Hong Kong, Peking und teilweise Singapur und Shanghai die Vorherrschaft in der Region teilen. Übereinstimmend mit der GaWC-Forschungsgruppe (2011) und Wall und van der Knapp (2011) ist die überdurchschnittlich hohe Zentralität von Hong Kong in wirtschaftlicher Hinsicht zu beachten. Drittens, profiliert sich Paris in den vorliegenden Resultaten als dritte zentralste Stadt hinter London und New York. Die zentrale Stellung von Paris steht im Einklang mit den Erkenntnissen von Alderson und Beckfield (2004; Alderson et al. 2010) und Wall und van der Knapp (2011). Diese Forscher erachten die Zentralität von Paris als äquivalent zu derjenigen von London und New York. Obwohl sowohl London als auch Paris europäische Städte sind und ein ähnliches funktionales Profil aufweisen, unterscheiden sie sich hinsichtlich ihres *binterworld* (Taylor 2004). Das *binterworld* von Paris umfasst die kontinentaleuropäischen, die nordafrikanischen sowie die nahöstlichen und die südamerikanischen Städte. London profiliert sich durch ein sehr heterogenes und geographisch diffuses *binterworld*; asiatische, ozeanische, skandinavische, israelische Städte sowie die Städte der Sub-Sahara gehören dazu. Auf New York beziehen sich die nord- und zentralamerikanischen Städte.

Funktionen und funktionale Netze

Die funktionale Differenzierung ist Kern und Hauptbeitrag der vorliegenden Arbeit und füllt eine wichtige empirische Forschungslücke. Eine methodische Einschränkung hinsichtlich der Resultate ist jedoch zu erwähnen. Die funktionalen Städtenetze werden nur zwischen den 95 Hubs berechnet und nicht für das ganze Städtetz. Die Resultate beschränken sich somit auf die funktionsbedingte Zentralität der 95 Hubs und auf die Struktur des Netzwerks zwischen diesen Städten. Der Grund dafür ist die grosse Menge an Daten. Dank der methodischen Vorleistung dieser Arbeit, insbesondere der Methodevaluation, und mit schnelleren Rechnern wäre zukünftig die funktionale Erforschung des gesamten Netzwerks möglich.

Aus den Daten haben sich acht Funktionen herauskristallisiert, wobei zwei davon sich in zwei bzw. drei Subfunktionen teilen. Es handelt sich um WIRTSCHAFT/TECHNOLOGIE, mit drei Subfunktionen COMPUTING, TECHNOLOGIE und WIRTSCHAFT, POLITIK, PERSONEN, GEBaute/NATÜRLICHE UMWELT, SPORT, KUNST, BILDUNG und GESELLSCHAFT/GESCHICHTE/MILITÄR mit den Subfunktionen MILITÄR/KONFLIKTE und GESCHICHTE/GESELLSCHAFT.

In absoluten Zahlen gesehen sind GEBaute/NATÜRLICHE UMWELT und KUNST die zwei wichtigsten Funktionen im Städtetz aus Wikipedia. Die Relevanz, bestimmt durch den Vergleich mit einem Zufallsdatensatz, ist hingegen für die Funktionen PERSONEN und GESELLSCHAFT/GESCHICHTE/MILITÄR und für die Subfunktion WIRTSCHAFT besonders hoch. Dies zeigt, dass insbesondere die personengebundenen und sozioökonomischen Funktionen die Integration der Städte ins Netzwerk bewirken.

Ein Vergleich der Resultate mit der Typologie der Funktionen nach Smith und Timberlake (1995a) (Tabelle 1) zeigt, dass die Resultate in dieser Arbeit eine höhere funktionale Auflösung aufweisen. Die Funktion WIRTSCHAFT/TECHNOLOGIE greift die Funktion Wirtschaft nach Smith und Timberlake (1995a) auf. Der Schwerpunkt liegt in beiden Definitionen bei den Unternehmen. In dieser Arbeit wird zudem die Technologie in die wirtschaftliche Funktion integriert. Die politische Funktion in dieser Arbeit spiegelt die politisch-institutionellen und rechtlichen Aspekte der Funktion Politik nach Smith und Timberlake (1995a) wider. Das Militär ist nach Smith und Timberlake (1995a) auch ein integraler Bestandteil der Funktion Politik. In der vorliegenden Arbeit ist es hingegen in der Funktion GESELLSCHAFT/GESCHICHTE/MILITÄR integriert. Diese bildet auch die Funktion Gesellschaft nach Smith und Timberlake (1995a) ab. Während nach Smith und Timberlake (1995a) zur gesellschaftlichen Funktion Familien und humanitäre Hilfe gehören, heben die Resultate den religiösen, philosophischen und historischen Aspekt hervor. Nach Smith und Timberlake (1995a) wird die Funktion Kultur inhaltlich breit definiert und umfasst die Funktionen KUNST und BILDUNG dieser Arbeit. Obwohl Smith und Timberlake (1995a) dem Sport keine funktionale Bedeutung beimessen, beeinflusst die Funktion Sport die Integration verschiedener Städte. Die Funktion GEBaute/NATÜRLICHE UMWELT wird von Smith und Timberlake (1995a) auch nicht erwähnt. Konzeptionell entspricht diese Funktion jedoch der ersten Ebene des „Raumes

der Ströme“ nach Castells (2001). Da in der Funktion GEBaute/NATÜRLICHE UMWELT Flughäfen und Flugzeuge eine zentrale Rolle spielen, ist diese Funktion auch kongruent zu den Flugpassagierbewegungsdaten.

Smith und Timberlake (1995a) unterscheiden Interaktionen nicht nur nach Funktion, sondern auch aufgrund deren Form (Tabelle 1). Diese Unterscheidung ist mit dem vorliegenden Ansatz nur beschränkt möglich. Einzig die Funktion PERSONEN in dieser Arbeit entspricht der Form Menschen nach Smith und Timberlake (1995a). Eine Unterscheidung zwischen Gütern und Informationen ist aufgrund der Daten nicht möglich.

Mit der funktional differenzierten Untersuchung der Integration der 95 Hubs sind teilweise ausgeprägte, funktionale Schwerpunkte feststellbar. So profiliert sich OST-/SÜDOSTASIEN als WIRTSCHAFTLICH/TECHNOLOGISCH orientierte Region. Nicht alle Städte der Region sind dabei gleichermassen von dieser Funktion geprägt. Die zentralsten Städte wie Hong Kong, Singapur und Shanghai haben einen ausgeprägten ökonomischen Charakter. Taylor (2005a) zählt hingegen Hong Kong, Singapur sowie Tokyo, aber nicht Shanghai zu den ökonomisch spezialisierten, führenden globalen Städte. In den vorliegenden Resultaten hat Tokyo jedoch eine überdurchschnittliche Zentralität in den Bereichen SPORT und COMPUTING. Die wirtschaftliche Zentralität ist hoch, jedoch nicht überdurchschnittlich. Zu den wichtigen sportlichen Ereignissen in Tokyo gehören die Sommerolympiade 1964, die geplante und wegen des zweiten Weltkriegs abgesagte Sommerolympiade 1940, sowie die Kandidaturen für die Olympiaden 2016 und 2020. Die multifunktionale Einbettung Tokyos ist derjenigen von Peking und Melbourne sehr ähnlich. Die starke WIRTSCHAFTLICHE Prägung von Hong Kong führt dazu, dass das WIRTSCHAFTLICHE Netzwerk vier Hauptzentren hat; neben London, Paris und New York etabliert sich Hong Kong als viertes Zentrum.

Die EUROPÄISCHEN Städte weisen eine unterdurchschnittliche WIRTSCHAFTLICH/TECHNOLOGISCHE Zentralität auf. Die Stärke dieser Städte liegt eher bei den Funktionen PERSONEN, GESELLSCHAFT/GESCHICHTE/MILITÄR und KUNST. Dies überrascht wegen des historischen Hintergrunds der europäischen Städte, sowohl aus kunsthistorischer Perspektive, als auch hinsichtlich der langen Kriegzeiten, nicht wirklich. Besonders ausgeprägt sind dabei die Zentralitätswerte für Rom und Wien. Es ist nicht erstaunlich, dass die GESELLSCHAFT/GESCHICHTE/MILITÄR Funktion auch eine sehr wichtige Rolle für die nahöstlichen und nordafrikanischen Städte spielt. Insbesondere Jerusalem, aber auch Alexandria und Damaskus stechen dabei hervor.

Die NORDAMERIKANISCHEN Städte sind in funktionaler Hinsicht den EUROPÄISCHEN Städten ähnlich, weisen jedoch einen schwächeren SOZIO/HISTORISCH/MILITÄRISCHEN Einfluss auf, dafür spielt die wirtschaftliche Dimension eine wichtigere Rolle. Die ZENTRAL-/SÜDAMERIKANISCHEN Städte weisen eine starke Prägung durch die Funktion SPORT auf. In Anbetracht der zentralen gesellschaftlichen Stellung des Fussballsports in Süd- und Zentralamerika, was sich z.B. in der Austragung von sechs Fussballweltmeisterschaften zwischen 1930 und 2014 und in den Siegen der süd- und zentralamerikani-

schen Mannschaften an solchen Sportanlässen zeigt, erstaunt die Stärke dieser Funktion nicht.

Eine bemerkenswerte Erkenntnis ergibt sich in Bezug auf die Funktion NATÜRLICHE/GEBAUTE UMWELT: für keine der zentralen Städte hat diese Funktion eine überdurchschnittliche Bedeutung. Dies könnte als Bestätigung des Postulats von Taylor (2001) gelten, gemäss dem die Infrastruktur zwar eine wichtige Voraussetzung für die Integration der Städte ins globale Netzwerk sind, jedoch nicht deren Zentralität bestimmt.

Die diskutierten Resultate stimmen nur punktuell mit den einzigen, bisherigen multifunktionalen Analysen von Taylor (2005a) überein. Neben der Kongruenz von Hong Kong und Singapur in wirtschaftlicher Hinsicht, stimmt nur die Stellung von Paris, London und New York überein. In beiden Analysen kristallisieren sich diese Städte als funktionsübergreifende globale Städte heraus. Diese Inkongruenz ist massgeblich von der sehr unterschiedlichen Operationalisierung der Funktionen bestimmt. So wird die Funktion Gesellschaft von Taylor (2005a) anhand von wissenschaftlichen Artikeln und NGOs operationalisiert. Gemäss den vorliegenden identifizierten Funktionen gehören die NGOs zur Funktion POLITIK und die wissenschaftlichen Publikationen zur Funktion BILDUNG.

Aus der obigen Diskussion über das Städtetz sind insbesondere drei Erkenntnisse hervorzuheben:

- (WEST-)EUROPA und NORDAMERIKA auf regionaler Ebene sowie London, Paris und New York bilden den Kern des globalen Städtetzes. Die Geographie der globalen Städte folgt ansatzweise dem Zentrum-Peripherie-Modell.
- Die OST-/SÜDOSTASIATISCHEN Städte entwickeln sich zu aufstrebenden globalen Städten mit einem ökonomischen Schwerpunkt. Jedoch ist nicht Tokyo alleine die zentralste Stadt, vielmehr entwickelt sich eine Hierarchie mit einer flachen Spitze, die von Tokyo, Hong Kong, Singapur, Peking und Shanghai besetzt wird.
- Aus geographischer Sicht sind die neuen Erkenntnisse über die AFRIKANISCH/NAHÖSTLICHEN Städte hervorzuheben in dieser Arbeit. Diese sind peripher und haben einen stärkeren Bezug zu einzelnen Städten der europäischen Region als zueinander, was zu einer hohen inneren Fragmentierung dieser Region führt.

9 Schlussfolgerungen

Die vorliegende Arbeit ist mit dem Ziel entstanden, neue Datengrundlagen, insbesondere nutzergenerierte Inhalte aus dem Web 2.0 zu erschliessen und einen theoriegeleiteten methodischen Rahmen zu entwickeln, mit dem sich ein besseres Verständnis von Städten als Knoten eines globalen Städtetzes gewinnen lässt. Die folgenden Abschnitte fassen die wichtigsten Erkenntnisse hinsichtlich der drei Pfeiler der Arbeit, nutzergenerierte Datengrundlage, interdisziplinärer methodischer Rahmen, sowie multifunktionale Erkenntnisse über das globale Städtetz, zusammen. Im letzten Teil werden als Ausblick die relevantesten Stossrichtungen für die zukünftige Forschung aufgezeigt.

Fazit

Nutzergenerierte Datengrundlage. Die online Enzyklopädie Wikipedia in englischer Sprache wurde als neue Datengrundlage für die Analyse des globalen Städtetzes evaluiert. Wikipedia entspricht den gestellten Anforderungen und ist somit eine Antwort auf Short's (1996) „*dirty little secret*“ im Forschungsbereich der globalen Städte. Wikipedia hat sich insbesondere dank ihrer relationalen, globalen und funktional differenzierten Struktur als Datengrundlage für die Untersuchung des globalen Städtetzes bewährt. Damit kann die Lücke zwischen den komplexen und vielfältigen theoretischen Ansätzen und der bisher geringen Zahl an empirischen Umsetzungen (Taylor 1997, 2004) geschlossen werden. Die von Denning et al. (2005) und Flanagin und Metzger (2008) hervorgehobenen Risiken bei der wissenschaftlichen Erforschung und Nutzung von Wikipedia wurden hinsichtlich ihres Einflusses auf die vorliegende Arbeit systematisch evaluiert. Erstens beeinträchtigt die Volatilität der Einträge, die durch die laufenden Aktualisierungen

verursacht wird, die Robustheit der Datengrundlage nicht. Zweitens lässt sich zwar eine räumliche Verzerrung, insbesondere ein leichtes Übergewicht der US-amerikanischen Städte identifizieren, aber dies ist mit entsprechender methodischer Vorsicht korrigierbar.

Wikipedia ist eines der umfangreichsten *crowdsourcing*-Projekte und wegen des geographischen Inhaltes auch von besonderer Relevanz für den Forschungsbereich der „*volunteered geography*“ (Goodchild 2007). In dieser Hinsicht rundet die vorliegende Studie die Arbeit von Hecht (2007) durch eine konkrete Fallstudie, in der eine explizite und thematisch konkrete wirtschaftsgeographische Forschungsfrage beantwortet wird, ab.

Interdisziplinärer methodischer Rahmen. Mit vorliegender Arbeit wurde ein interdisziplinärer methodischer Rahmen, der sich auf eine Triangulation von Netzwerkanalyse und –visualisierung, semantische Analyse und Methoden der *spatialization* stützt, entwickelt. Der visuell-analytische methodische Rahmen eignet sich für Netzwerkdaten aus unterschiedlichen Themenbereichen, wie etwa die Wissensvisualisierung (Fabrikant und Salvini 2011). Der methodische Rahmen ist theoriegeleitet und greift klassische Aspekte aus der langen kartographischen und netzwerkanalytischen Tradition auf und kombiniert sie mit neuen, skalierbaren methodischen Ansätzen aus der Geographischen Informationswissenschaft und der Informatik. Der umfassende methodische Rahmen wurde mit der Analyse des globalen Städtetzes getestet. Damit werden bisherige methodische Ansätze im Bereich der Erforschung des globalen Städtetzes um einen formalen und systematischen visuell-analytischen Ansatz ergänzt.

Die Methodik umfasst drei Forschungsschritte: Modellierung, Analyse der Daten, sowie Evaluation der Ergebnisse. Hinsichtlich der Modellierung wurde die Transformation der Rohdaten aus Wikipedia zu einem funktional differenzierten Städtetz mittels computerbasierter semantischer Analyse umgesetzt und diskutiert. Im analytischen Teil der Arbeit wurden netzwerkanalytische und geoinformatische Methoden für die prägnante Darstellung der relevanten Eigenschaften des globalen Städtetzes präsentiert. Die Methodenevaluation hat sich neben der Robustheit der Datengrundlage mit der Genauigkeit der Verräumlichung sowie der Methodenabhängigkeit der eruierten Funktionen beschäftigt. Die Genauigkeit der Verräumlichung mit dem DrL/VxOrd-Algorithmus ist skalenabhängig und stark von der vorgängigen Datenfilterung bestimmt. Die acht thematischen Schwerpunkte der Funktionen (PERSONEN, GEBaute/NATÜRLICHE UMWELT, POLITIK, WIRTSCHAFT/TECHNOLOGIE, SPORT, BILDUNG, GESELLSCHAFT/GESCHICHTE/MILITÄR und KUNST) ergeben sich unabhängig von der gewählten Klassierungsmethode. Die kognitiv adäquate graphische Umsetzung der Resultate wurde nicht empirisch mit Probanden evaluiert, sondern stützt sich auf empirisch-basierte Gestaltungsempfehlungen von Fabrikant et al. (2004; 2006) und Montello et al. (2003).

Aus den drei thematischen Fallstudien lassen sich einerseits eine grundsätzliche Erkenntnis sowie verschiedene, spezifische inhaltliche Erkenntnisse herausarbeiten. Die relationalen Regionen sind alle von einer starken räumlichen Differenzierung gekenn-

zeichnet. Grundsätzlich zeigen alle Resultate eine ausgesprochene räumliche Prägung der relationalen Geographie. Die relationalen Regionen folgen mehrheitlich den subkontinentalen oder kontinentalen politischen Grenzen. Daraus lässt sich schliessen, dass trotz des technologischen Wandels und der fortschreitenden Globalisierung, globale urbane Prozesse nicht losgelöst vom Raum stattfinden. Aussagekräftige Ausnahmen sind trotzdem zu vermerken. Das exemplarischste Beispiel dafür ist die Integration von Lhasa (Hauptstadt von Tibet) in die WESTEUROPÄISCHE Region. Der Grund dafür liegt vermutlich in der solidarischen, politischen Einstellung Europas gegenüber dieser Region.

Aus der ersten Fallstudie ist zudem bekannt, dass im Gegensatz zu Sassen (1991), Friedmann (1986, 1995) und Alderson und Beckfield (2004) New York, London und Paris die zentralsten Städte im globalen Städtenetz sind. Tokyo rangiert unter den ersten zehn Plätzen und ist ein wichtiges Zentrum in der OST-/SÜDOSTASIATISCHEN Region. Die neue Geographie der globalen Städte durchbricht nicht, wie von Sassen (1991) postuliert, das Zentrum-Peripherie-Modell. Die Ergebnisse bestätigen die empirischen Erkenntnisse von Alderson und Beckfield (2004) und stehen im Widerspruch zu den theoretischen Ausführungen von Sassen (1991).

Multifunktionale Erkenntnisse über das globale Städtenetz. Die empirischen Resultate dieser Arbeit vertiefen die bisherigen Erkenntnisse von Taylor (2005a) über das funktional differenzierte globale Netzwerk. Während Taylor (2005a) seine Resultate auf sehr heterogene Datenquellen stützt, ergeben sich die vorliegenden Resultate aus einer einzigen sehr umfangreichen Datengrundlage. Die acht Funktionen (PERSONEN, GEBÄUDE/NATÜRLICHE UMWELT, POLITIK, WIRTSCHAFT/TECHNOLOGIE, SPORT, BILDUNG, GESELLSCHAFT/GESCHICHTE/MILITÄR und KUNST) greifen die vier Funktionen (Wirtschaft, Politik, Kultur und Gesellschaft) nach Smith und Timberlake (1995a) auf, organisieren die thematischen Schwerpunkte aber umfassender und ergänzen sie insbesondere durch eine neue technologische, sportliche und infrastrukturelle Komponente. Die Funktion GEBÄUDE/NATÜRLICHE UMWELT greift die erste Ebene des Raumes der Ströme nach Castells (2001) wieder auf. Die modellierten funktionalen Städtenetze zeigen, dass die Hierarchie der globalen Städte über die Funktionen mehrheitlich stabil ist. Trotzdem lassen sich regionale und funktionale Schwerpunkte differenzieren. Die EUROPÄISCHEN Städte weisen eine überdurchschnittliche SOZIO-HISTORISCH-MILITÄRISCHE sowie KÜNSTLERISCHE Zentralität auf. Die OST-/SÜDOSTASIATISCHEN Städte, insbesondere Hong Kong profilieren sich in der Funktion WIRTSCHAFT/TECHNOLOGIE. Die NORDAMERIKANISCHEN Städte sind den EUROPÄISCHEN ähnlich, sind jedoch weniger stark SOZIO-HISTORISCH-MILITÄRISCH beeinflusst, dafür spielt die WIRTSCHAFTLICH/TECHNOLOGISCHE Funktion eine wichtige Rolle. Die SÜD-/ZENTRALAMERIKANISCHEN Städte sind stärker von der Funktion SPORT und die nahöstlichen und nordafrikanischen Städte von der SOZIO-HISTORISCH-MILITÄRISCHEN Funktion geprägt.

Die vorliegende Arbeit füllt eine Lücke im Forschungsbereich der globalen Städte. Im Spezifischen werden erstens, eine neue nutzergenerierte Datenquelle erschlossen, zweitens, ein interdisziplinärer methodischer Ansatz entwickelt und drittens, neue Erkenntnisse über das multifunktionale globale Städtenetz generiert. Aufbauend auf den gewonnenen Erkenntnissen lassen sich sowohl weiterführende methodische als auch thematische Forschungsschwerpunkte entwickeln. Diese werden im nächsten Abschnitt vorgelegt.

Ausblick

In diesem Abschnitt werden vier Anknüpfungspunkte für die zukünftige Forschung diskutiert. Zwei davon beziehen sich auf die Methodik und betreffen die Evaluation der Visualisierungen und die Weiterentwicklung derselben mit Animationen. Auf thematischer Ebene beziehen sich die übrigen Punkte auf die zeitliche Dimension, die räumliche Massstabsebene sowie die funktionalen Netzwerke.

Wie bereits erwähnt fokussiert die Evaluation in dieser Arbeit auf die Robustheit der Datengrundlagen sowie auf die Validität der Methoden. Weiterführende empirische Probandenstudien zur Wahrnehmung der präsentierten Darstellungen (z.B. Abbildung 33, oder Abbildung 39) sollten in Anlehnung an die Evaluation von einfachen Netzwerkdarstellungen in 2D (Fabrikant et al. 2004), in 3D (Maggi 2009), oder flächenhaften Darstellungen (Fabrikant et al. 2006), durchgeführt werden. Die Kombination solcher Evaluationsmethoden mit Augenbewegungsanalysen könnte zudem neue Erkenntnisse über den komplexen Lesevorgang von Netzwerkdarstellungen hervorbringen.

Eine zweite Vertiefungsmöglichkeit im Zusammenhang mit der graphischen Darstellung ist die Entwicklung von animierten Visualisierungen, um der zeitlichen Dimension der Daten gerecht zu werden (DiBiase et al. 1992). So könnten die zeitbedingten, aber auch die funktionsbedingten Änderungen des Netzwerkes und ein Wechsel der Stellungen einzelner Städte im Netzwerk visuell mit Animationen umgesetzt werden. Zu beachten ist jedoch, dass nach Slocum et al. (2004) der Nutzen von Animation aufgabeabhängig ist. Griffin et al. (2006) bestätigen mit empirischen Studien die Wirksamkeit von Animation in der Erkennung von raum-zeitlichen Clustern, räumen jedoch ein, dass eine vertieftere Untersuchung, insbesondere die Bestimmung der kognitiven Last von Animationen, notwendig ist. Entsprechende empirische Evaluationen wären somit auch in dieser Hinsicht unentbehrlich.

Thematisch beschränkt sich die vorliegende Arbeit auf drei ausgewählte Fallstudien. Es bieten sich dementsprechend verschiedene Vertiefungsmöglichkeiten an. Die Struktur und Hierarchie der einzelnen Regionen könnte detaillierter analysiert und somit miteinander verglichen werden. Insbesondere wäre eine genauere Auseinandersetzung mit den bislang wenig erforschten afrikanischen Städten naheliegend. Fünf Aspekte sollten an dieser Stelle jedoch besonders hervorgehoben werden:

1. Die funktionalen Netzwerke wurden mit 95 Hubs berechnet. Dank den neu gewonnenen Erkenntnissen über das Potential dieser nutzergenerierten Daten und

über die Robustheit der Methoden lässt sich der Ansatz, die notwendige Rechenkapazität vorausgesetzt, auf das gesamte Netzwerk zwischen allen 3'532 Städten übertagen.

2. Der MST als visuell-analytische Methode hat den Vorteil, die prägende, latente hierarchische Struktur des globalen Städtenetzes hervorzuheben. Da sowohl das globale Städtenetz, als auch die Einbettung jeder einzelnen Stadt in dieses Netzwerk sehr komplex ist, sind jedoch neben dieser zentralen hierarchischen Eigenschaft noch weitere sekundäre Muster zu erwarten. Eine ergänzende Analyse der Gesamtheit der intraregionalen Beziehungen oder des Egonetzwerks (Netzwerk mit allen Beziehungen einer ausgewählten Stadt) jeder einzelnen Stadt würde diese Muster zur Erscheinung bringen und dadurch zusätzliche Erkenntnisse über die Komplexität im globalen Städtenetz liefern.
3. Die Diskussion der Resultate hat ergeben, dass das Potential von Wikipedia als Datengrundlage auch in Bezug auf die räumliche Massstabebene noch nicht ausgeschöpft ist. Die Erforschung von kontinentalen, nationalen und regionalen Städtenetzen unter Betrachtung einer grossen Anzahl Städte wäre also durchaus angebracht. Insbesondere die Erforschung von *global city regions* (Scott 2001) könnte vom vorgeschlagenen Ansatz besonders profitieren. Die Anwendung des entwickelten Ansatzes auf die Region Ostschweiz mit der deutschsprachigen Version von Wikipedia als Datenbasis hat sich bereits als vielversprechend erwiesen, um lokale Strukturen und Organisation zu erschliessen (Bruggmann 2012). Bei lokalen Analysen ist insbesondere die Datengrundlage hinsichtlich ihrer Robustheit genau zu evaluieren. Die Anzahl Links pro Stadt-Artikel könnte zu gering sein um aussagekräftige Resultate zu generieren.
4. Ein letzter Anknüpfungspunkt ist der Einbezug der zeitlichen Dimension. Wie in der Diskussion erwähnt umfasst Wikipedia als Datengrundlage nicht nur eine Momentaufnahme, sondern schliesst auch Ereignisse der Zukunft und der Vergangenheit, die zum gegenwärtigen Zeitpunkt von Relevanz sind, mit ein. In dieser Hinsicht wäre neben der funktionalen auch eine zeitliche Differenzierung des Städtenetzes und somit die Erforschung der raum-zeitlichen Veränderungen möglich und erstrebenswert. Die Akteur-Artikel lassen sich aufgrund zeitlicher Kategorien (z.B. „2002 films“, „2009 in France“, „2000“, „19th century“) zeitlich kodieren und somit differenzieren. Im vorliegenden Datensatz lassen sich etwa zwei Drittel der Akteur-Artikel einem bestimmten Jahr zuweisen. Die zeitliche Abdeckung müsste aber sorgfältig untersucht werden. Insbesondere zeigt sich erstens eine tiefere Anzahl Artikel betreffend Themen, die vor Lancierung von Wikipedia (2001) stattgefunden haben, zweitens nimmt die Anzahl nach 2001 kontinuierlich zu und drittens ist bei allen Kategorien die zeitliche Auflösung nicht klar definiert. Kategorien mit der Beschreibung 1990 beziehen sich sowohl auf das Jahr 1990, als auch auf das Jahrzehnt der 90er Jahre. Trotzdem ist die Er-

forschung der zeitlichen Dimension insbesondere in Zukunft, wenn sich der Inhalt auch zeitlich stabilisiert, erstrebenswert und vermutlich auch möglich.

10 Bibliographie

- ALDERSON, A.S. und BECKFIELD, J. (2004): Power and position in the world city system. *American Journal of Sociology*. 109(4): 811-851.
- ALDERSON, A.S., BECKFIELD, J. und SPRAGUE-JONES, J. (2010): Intercity relations and globalisation. The evolution of the global urban hierarchy, 1981-2007. *Urban Studies*. 47(9): 1899-1923.
- ANDRIEU, D., KAISER, C. und OUREDNIK, A. (2008): *Scape Toad 1.1*.
- BEAVERSTOCK, J.V., SMITH, R.G. und TAYLOR, P.J. (2000a): World-city network. A new metageography? *Annals of the Association of American Geographers*. 90(1): 123-134.
- BEAVERSTOCK, J.V., SMITH, R.G., TAYLOR, P.J., WALKER, D.R.F. und LORIMER, H. (2000b): Globalization and world cities. some measurement methodologies. *Applied Geography*. 20(1): 43-63.
- BENNETT, C., RAYALL, J., SPALTEHOLZ, L. und GOOCH, A. (2007): The aesthtics of graph visualisation. *International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, Banff.
- BERRY, B. (1964): Cities as systems within systems of cities. *Papers and proceedings of the regional science association*. 13(1): 147-163.
- BERTIN, J. (1974): *Graphische Semiologie. Diagramme Netze Karte*. Berlin, New York.

- BLONDEL, V., GUILLAUME, J.-L., LAMBIOTTE, R. und LEFEBRE, E. (2008): Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. -(10).
- BORGATTI, S.P. und EVERETT, M.G. (2006): A graph-theoretic perspective on centrality. *Social networks*. 28(4): 466-484.
- BORGATTI, S.P., EVERETT, M.G. und FREEMAN, L.C. (2002): *UCINET 6 for Windows. Software for social network analysis*.
- BÖRNER, K., CHEN, C. und BOYACK, K. (2003): Visualizing knowledge domains. *Annual review of information science and technology*. 37(179-255).
- BOYACK, K.W., KLAVANS, R. und BÖRNER, K. (2005): Mapping the backbone of science. *Scientometrics*. 64(3): 351-374.
- BRANDES, U. (2001a): Drawing on physical analogies. In: KAUFMANN, M. und WAGNER, D. (eds.): *Drawing graphs. Methods and models*. Berlin: 71-86.
- BRANDES, U. (2001b): A fast algorithm for betweenness centrality. *Journal of mathematical sociology*. 25(2): 163-177.
- BRASS, D.J. (1995): A social network perspective on human resources management. *Research in personnel and human resources management*. 13(1): 39-79.
- BRENNER, N. und KEIL, R. (eds.) (2006): *The global cities reader*. New York.
- BREWER, C.A. (2005): *Designing better maps. A guide for GIS users*. Redlands.
- BREWER, C.A. (2006): *ColorBrewer 2.0*. Aus: <http://colorbrewer2.org/>, Zugriff: 14.11.2011.
- BRUGGMANN, A. (2012): *Netzwerkvisualisierung der Ostschweiz. Die Raumgliederung der Schweiz mit Wikipedia neu formuliert*. Masterarbeit, Geographisches Institut, Universität Zürich.
- BUDANITSKY, A. und HIRST, G. (2006): Evaluating WordNet-based measures of lexical semantic relatedness. *Computational linguistics*. 32(1): 13-47.
- BURGER, M.J., WALL, R. und VAN DER KNAAP, B. (2008): Measuring urban competition on the basis of flows between cities. Some evidence from the world city network. *GaWC Research Bulletin*. 273.
- BUTTENFIELD, B.P. und MCMASTER, R.B. (eds.) (1991): *Map generalization. Making rules for knowledge representation*. London.
- CARD, S.K., MACKINLAY, J.D. und SCHNEIDERMAN, B. (eds.) (1999): *Readings in information visualization. Using vision to think*. San Francisco.
- CASTELLS, M. (1996): *The rise of the network society*. New York.
- CASTELLS, M. (2001): *Das Informationszeitalter I. Der Aufstieg der Netzwerkgesellschaft*. Opladen.
- CASTELLS, M. (2002a): *Das Informationszeitalter II. Die Macht der Identität*. Opladen.

- CASTELLS, M. (2002b): Local and global. cities in the network society. *Tijdschrift voor Economische en Sociale Geografie*. 93(5): 548-558.
- CASTELLS, M. (2003): *Das Informationszeitalter III. Jahrtausendwende*. Opladen.
- CASTELLS, M. (2009): *Communication power*. Oxford.
- CASTELLS, M. (2010): Globalisation, networking, urbanisation. reflection on the spatial dynamics of the informational age. *Urban Studies*. 47(13): 2737-2745.
- CHEN, C. (1999): Visualising semantic space and author co-citation networks in digital libraries. *Information processing and management*. 35(3): 401-420.
- CHEN, C. (2003): *Mapping scientific frontiers. The quest of knowledge visualization*. Hingham.
- CHEN, C. (2004): *Information visualization. Beyond the horizon*. London.
- COHEN, J. (1960): A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20(1): 37-46.
- COHEN, R.B. (1981): The new international division of labour, multinational corporations und urban hierarchy. In: DEAR, M. und SCOTT, A.J. (eds.): *Urbanization and urban planning in capitalist society*. New York: 287-315.
- COLEMAN, M.K. und PARKER, D.S. (1996): Aesthetics-based graph layout for human consumption. *Software - Practice and experience*. 26(12): 1415-1438.
- COUCLELIS, H. (1992): Location, place, region, and space. In: ABLER, R., et al. (eds.): *Geography's inner worlds: pervasive themes in contemporary American geography* New Brunswick: 215-233.
- COUCLELIS, H. (1998): Worlds of information. The geographic metaphor in the visualization of complex information. *Cartography and Geographic Information Science*. 25(4): 209-220.
- DEARHOLT, D.W. und SCHVANENVELDT, R.W. (1990): Properties of pathfinder networks. In: SCHVANENVELDT, R.W. (eds.): *Pathfinder associative networks: studies in knowledge organization*. Norwood: 1-30.
- DENNING, P., HORNING, J., PARNAS, D. und WEINSTEIN, L. (2005): Wikipedia risks. *Communications of the ACM*. 48(12).
- DERUDDER, B. (2006): On conceptual confusion in empirical analyses of a transnational urban network. *Urban Studies*. 43(11): 2027-2046.
- DERUDDER, B., TAYLOR, P.A., NI, P., DE VOS, A., HOYLER, M., HANSSENS, H., BASSENS, D., HUANG, J., WITLOX, F., SHEN, W. und YANG, X. (2010): Pathways of change. Shifting connectivities in the world city network, 2000-08. *Urban Studies*. 47(9): 1861-1878.
- DERUDDER, B. und TAYLOR, P.J. (2005): The cliquishness of world cities. *Global Networks*. 5(1): 71-91.

- DERUDDER, B., TAYLOR, P.J., WITLOX, F. und CATALANO, G. (2003): Hierarchical tendencies and regional pattern in the world city network. A global urban analysis of 234 cities. *Regional Studies*. 37(9): 875-886.
- DERUDDER, B. und WITLOX, F. (2004): Assessing central places in a global age. On the networked localization strategies of advanced producer services. *Journal of Retailing and Consumer Services*. 11(3): 171-180.
- DERUDDER, B. und WITLOX, F. (2008): Mapping world city network through airline flows. context, relevance, and problems. *Journal of Transport Geography*. 16(5): 305-312.
- DI BATTISTA, G., EADES, P., TAMASSIA, R. und TOLLIS, I.G. (1994): Algorithms for drawing graphs. an annotated bibliography. *Computational geometry: Theory and applications*. 4(5): 235-282.
- DIBIASE, D., MACEACHREN, A.M., KRYGIER, J.B. und REEVES, A. (1992): Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Systems*. 19(4): 201-214.
- DIENER, R., HERZOG, J., MEILI, M., DE MEURON, P. und SCHMIDT, C. (2006): *Die Schweiz ein städtebauliches Portrait*. Basel.
- EADES, P. (1984): A heuristic for graph drawing. *Congressus Numerantium*. 42(-): 149-160.
- ELWOOD, S. (2010): Geographic information science. emerging research on the societal implications of the geospatial web. *Progress in Human Geography*. 34(3): 349-357.
- ESRI (2011): *ArcGIS desktop 10*.
- ESULI, A. und SEBASTIANI, F. (2006): SentiWordNet: A publicly available lexical resource for opinion mining. *Proceedings of LREC*. 6(-): 417-422.
- FABRIKANT, S.I. (2000): *Spatial metaphors for browsing large data archives*. PhD Thesis, Departement of geography, University of Colorado.
- FABRIKANT, S.I. (2001a): Evaluating the usability of the scale metaphor for querying semantic spaces. *Spatial Information Theory, International Conference (COSIT)*, Morro Bay.
- FABRIKANT, S.I. (2001b): Visualizing region and scale in information spaces. *International Cartographic Conference (ICC)*, Beijing.
- FABRIKANT, S.I. (2004): Abstraction and scale in spatialization. *GeoCart'2003, National Cartographic Conference*, Taupo.
- FABRIKANT, S.I. und BUTTENFIELD, B.P. (2001): Formalizing semantic spaces for information access. *Annals of the Association of American Geographers*. 91(2): 263-280.
- FABRIKANT, S.I., MONTELLO, D., RUOCCO, M. und MIDDLETON, R. (2004): The distance-similarity metaphor in network display spatializations. *Cartography and Geographic Information Science*. 31(4): 237-252.

- FABRIKANT, S.I., MONTELLO, D.R. und MARK, D.M. (2006): The distance similarity metaphor in region-display spatialization. *IEEE Computer Graphics and Applications*. 26(4): 34-44.
- FABRIKANT, S.I. und SALVINI, M.M. (2011): Charting the ICA world of cartography 1999–2009. *International Cartographic Conference (ICC)*, Paris.
- FABRIKANT, S.I. und SKUPIN, A. (2005): Cognitively plausible information visualization. In: DYKES, J., et al. (eds.): *Exploring Geovisualization*. Amsterdam: 667-690.
- FLANAGIN, A.J. und METZGER, M.J. (2008): The credibility of vounteered geographic information. *GeoJournal*. 72(-): 137-148.
- FLEISCHER, R. und HIRSCH, C. (2001): Graph drawing and its applications. In: KAUFMANN, M. und WAGNER, D. (eds.): *Drawing graphs. Methods and models*. Berlin: 1-22.
- FORD, L.R. und FULKERSON, D.R. (1956): Maximal flow through a network. *Canadian journal of mathematics*. 8(-): 399-404.
- FREEMAN, L.C. (1979): Centrality in social networks. I. conceptual clarification. *Social networks*. 1(3): 215-239.
- FRICK, A., LUDWIG, A. und MEHLDAU, H. (1994): A fast adaptive layout algorithm for undirected graphs. *DIMACS International Workshop on Graph Drawing*, Princeton.
- FRIEDMANN, J. (1986): The world city hypothesis. *Development and change*. 17(1): 69-84.
- FRIEDMANN, J. (1995): Where we stand. A decade of world ctiy research. In: KNOX, P.L. und TAYLOR, P.J. (eds.): *World Cities in a World System*. Cambridge, New York: 21-47.
- FRIEDMANN, J. und WOLFF, G. (1982): World city formation. An agenda for research and action. *International Journal of Urban and Regional Research*. 6(3): 309-344.
- FRUCHTERMAN, T.M.J. und REINGOLD, E.M. (1991): Graph drawing by force-directed placement. *Software - Practice and experience*. 21(11): 1129-1164.
- GABRILOVICH, E. und MARKOVITCH, S. (2007): Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *International Joint Conference on Artificial Intelligence*, Hyderabad.
- GARLANDINI, S. und FABRIKANT, S.I. (2009): Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. *Spatial Information Theory, International Conference (COSIT)*, Belfast.
- GAWC (2009a): *The world according to GaWC 2000* Aus: <http://www.lboro.ac.uk/gawc/world2000t.html>, Zugriff: 16.04.2012.
- GAWC (2009b): *The world according to GaWC 2004*. Aus: <http://www.lboro.ac.uk/gawc/world2004t.html>, Zugriff: 16.04.2012.
- GAWC (2010): *The world according to GaWC 2008*. Aus: <http://www.lboro.ac.uk/gawc/world2008t.html>, Zugriff: 16.04.2012.

- GAWC (2011): *The world according to GaWC 2010*. Aus: <http://www.lboro.ac.uk/gawc/world2010t.html>, Zugriff: 16.04.2012.
- GAWC (2012): *What we are about. Mission statements*. Aus: <http://www.lboro.ac.uk/gawc/mission.html>, Zugriff: 02.04.2012.
- GEDDES, P. (1915): *Cities in evolution. An introduction to the town planning movement and to the study of civics*. London.
- GILES, J. (2005): Internet encyclopaedias go head to head. *Nature*. 438(15): 900-901.
- GLOTT, R., SCHMIDT, P. und GHOSH, R. (2010): *Wikipedia survey - Overview of results*. Aus: http://www.wikipediaurvey.org/docs/Wikipedia_Overview_15March2010-FINAL.pdf, Zugriff: 03.03.2012.
- GNOS, A. (2010): *Geovisualisierung von Netzwerken*. Masterarbeit, Geographisches Institut, Universität Zürich.
- GOLLEDGE, R.G. (1995): Primitives of spatial knowledge. In: NYERGES, T.L., et al. (eds.): *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*. Dordrecht: 29-44.
- GOMORY, R.E. und HU, T.C. (1964): Synthesis of a communication network. *Journal of the society for industrial and applied mathematics*. 12(2): 348-369.
- GOODCHILD, M.F. (2007): Citizens as sensors. The world of volunteered geography. *GeoJournal*. 69(4): 211-221.
- GRIFFIN, A.L., MACEACHREN, A.M., HARDISTY, F., STEINER, E. und BONAN, L. (2006): A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*. 96(4): 740-753.
- GROSSMAN, L. (2006): Time's person of the year: You. In *Time*. New York.
- GUIMERÀ, R. und AMARAL, L.A.N. (2005a): Cartography of complex networks. Modules and universal roles. *Journal of Statistical Mechanics*. 1-12.
- GUIMERÀ, R. und AMARAL, L.A.N. (2005b): Functional cartography of complex metabolic networks. *Nature*. 433(7028): 895-900.
- GUIMERÀ, R., MOSSA, S., TURTSCHI, A. und AMARAL, L.A.N. (2005): The worldwide air transportation network. Anomalous centrality, community structure, and cities' global roles. *PNAS*. 102(22): 7794-7799.
- HECHT, B. (2007): *Utilizing Wikipedia as spatiotemporal knowledge repository*. Master Thesis, University of California (UCSB).
- HECHT, B. und GERGLE, D. (2010): The tower of babel meets Web 2.0. User-generated content and its application in a multilingual context. *Conferenz on human factors in computing systems (ACM)*, Atlanta.

- HECHT, B. und MOXLEY, E. (2009): Terabytes of Tobler. Evaluating the first law in a massive, domain-neutral representation of world knowledge. *Spatial Information Theory, International Conference (COSIT)*, Aber Wrac'h.
- HECHT, B. und RAUBAL, M. (2008): GeoSR. Geographically explore semantic relations in world knowledge. *International Conference on Geographic Information Science (AGILE)*, Girona.
- HOLLOWAY, T., BOZICEVIC, M. und BÖRNER, K. (2007): Analysing and visualizing the semantic coverage of Wikipedia and its authors. *Complexity*. 12(3): 30-40.
- HOWE, J. (2006): The rise of crowdsourcing. *Wired*.
- HYMER, S. (1972): The multinational corporation and the law of the uneven development. In: BHAGWATI, J.N. (eds.): *Economics and World Order. From the 1970's to the 1990's*. London.
- JOHNSON, M. (1987): *The body in the mind. The bodily basis of meaning, imagination, and reason*. Chicago.
- JONES, C.B. und PURVES, R.S. (2008): Geographical information retrieval. *International Journal of Geographic Information Science*. 22(3): 219-228.
- KAMADA, T. und KAWAI, S. (1989): An algorithm for drawing general undirected graphs. *Information Processing Letters*. 31(1): 7-15.
- KITTUR, A., CHI, E.H. und SUH, B. (2009): What's in Wikipedia? Mapping topics and conflict using socially annotated category structure. *International conference on Human factors in computing systems*, New York.
- KLAVANS, R. und BOYACK, K. (2006a): Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*. 57(2): 251-263.
- KLAVANS, R. und BOYACK, K. (2006b): Quantitative evaluation of large maps of science. *Scientometrics*. 68(3): 475-499.
- KLAVANS, R. und BOYACK, K. (2011): Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology*. 62(1): 1-18.
- KNOX, P.L. (1998): Globalization and world city formation. In: GRAVESTEIJN, S.J.E., et al. (eds.): *Timing global cities*. Utrecht: 21-31.
- KOFFKA, K. (1935): *Principles on gestalt psychology*. New York.
- KOHONEN, T. (2001): *Self-organizing maps*. Berlin.
- KORFF, R. (1987): The world city hypothesis. A critique. *Development and change*. 18(3): 483-495.
- KOS, A.J. und PSENICKA, C. (2000): Measuring cluster similarity across methods. *Psychological Reports*. 86(3): 858-862.

- LABRI (2010): *Tulip 3.5*.
- LACAYO-EMERI, M. (2011a): *An integrated toolset for exploration of spatiotemporal data using self-organizing maps and gis*. Departement of geography, San Diego State University.
- LACAYO-EMERI, M. (2011b): *SOM Analyst*. Aus: <http://code.google.com/p/somanalyst/>, Zugriff: 19.09.2011.
- LACAYO-EMERY, M. (2011): *An integrated toolset for exploration of spatiotemporal data using self-organizing maps and gis*. Master Thesis, Departement of geography, San Diego State University.
- LAKOFF, G. (1987): *Women, fire, and dangerous things. What categories reveal about the mind*. Chicago.
- LAKOFF, G. und JOHNSON, M. (1980): *Metaphors we live by*. Chicago.
- LYNCH, K. (1962): *The image of the city*. Boston.
- MACEACHREN, A.M. (1994): *Some truth with maps. A primer on design and symbolization*. Washington.
- MACEACHREN, A.M. (1995): *How maps work. Representation, visualization, and design* New York.
- MAGGI, S. (2009): *Evaluation of 3D spatializations*. Masterarbeit, Geographisches Insitut Universität Zürich.
- MANNING, C.D., RAGHAVAN, P. und SCHÜTZE, H. (2008): *Introduction to information retrieval. Classical and web information retrieval systems. algorithms, mathematical foundations and practical issues*. Cambridge.
- MARCUSE, P. und VAN KEMPEN, R. (2000): *Globalicing cities*. Oxford.
- MARK, D.M. (2003): Geographic information science. defining the field. In: DUCKHAM, M., et al. (eds.): *Foundations of Geographic Information Science*. New York: 1-15.
- MCMASTER, R.B. (1989): Introduction to "Numerical generalization in cartography". *Cartographica*. 26(1): 1-6.
- MONGE, P.R. und CONTRACTOR, N.S. (2003): *Theories of communication networks*. New York.
- MONTELLO, D., FABRIKANT, S., RUOCCO, M. und MIDDLETON, R. (2003): Testing the first law of cognitive geography on point-display spatializations. *Conference on Spatial Information Theory (COSIT)*, Ittingen, Schweiz.
- MORENO, J.L. (1933): *Who shall survive? Foundations of sociometry, group psychotherapy, and sociodrama*. New York.
- NATIONAL RESEARCH COUNCIL, R.G.C. (1997): *Rediscovering geography. New relevance for science and society*. Washington.

- NEAL, Z.P. (2012): Structural determinism in the interlocking world city network. *Geographical Analysis*. 44(2): 162-170.
- NEWMAN, M.E.J. (2004): Analysis of weighted networks. *Physics reviews E*. 70(5): 1-9.
- NEWMAN, M.E.J. und GIRVAN, M. (2004): Finding and evaluating community structure in networks. *Physics reviews E*. 59(-).
- NWB TEAM (2006): *Network Workbench Tool*.
- NWB TEAM (2011): *NWB community wiki*. Aus: <https://nwb.slis.indiana.edu/community/?n=Main.HomePage>, Zugriff: 07.09.2011.
- OPENSTREETMAP CONTRIBUTORS (2011a): *Stats*. Aus: <http://wiki.openstreetmap.org/wiki/Stats>, Zugriff: 22.06.2011.
- OPENSTREETMAP CONTRIBUTORS (2011b): *Was ist OpenStreetMap*. Aus: http://www.openstreetmap.ch/faq.php#was_ist_osm, Zugriff: 22.06.2011.
- PALMER, S.E. (1992): Common region. A new principle of perceptual grouping. *Cognitive psychology*. 24(3): 436-447.
- PALMER, S.E. und ROCK, I. (1994): Rethinking perceptual organization. The role of uniform connectedness. *Psychonomic bulletin and review*. 1(1): 29-55.
- PFLIEGER, G. und ROZENBLAT, C. (2010): Introduction. Urban networks and network theory. The city as the connector of multiple networks. *Urban Studies*. 47(13): 2723-2736.
- POMERANTZ, J.R. und SCHWARTZBERG, S.D. (1975): Grouping by proximity. Selective attention measures. *Perception and Psychophysics*. 18(5): 335-361.
- PONZETTO, S.P. und STRUBE, M. (2007): Deriving a large scale taxonomy from Wikipedia. *Journal of Artificial Intelligence Research*. 30(2): 181-212.
- PUMAIN, D. (2006a): Alternative explanations of hierarchical differentiation in urban systems. In: PUMAIN, D. (eds.): *Hierarchy in Natural and Social Sciences*. Dordrecht: 169-222.
- PUMAIN, D. (2006b): Systèmes de villes et niveaux d'organisation. In: BOURGINE, P. und LESNE, A. (eds.): *Morphogenèse*. Paris: 239-263.
- PUMAIN, D. und ROBIC, M.-C. (1996): Théoriser la ville. In: DERYCKE, P.-H., et al. (eds.): *Penser la ville: théories et modèles*. Paris: 107-161.
- PURCHASE, H.C. (1997): Which aesthetic has the greatest effect on human understanding. *International Symposium on Graph Drawing*, Rome.
- PURCHASE, H.C. (1998): Performance of layout algorithms. Comprehension, not computation. *Journal of visual languages and computing*. 9(6): 947-957.
- PURCHASE, H.C. (2000): Effective information visualisation. A study of graph drawing aesthetics and algorithms. *Interacting with computers*. 13(2): 147-162.

- PURCHASE, H.C., COHEN, R.F. und JAMES, M.I. (1997): An experimental study of the basic for graph drawing algorithms. *Journal of Experimental Algorithmics*. 2(4): 1-17.
- RAYMOND, E. (1997): The cathedral and the bazaar. *Knowledge, Technology & Policy*. 12(3): 23-49.
- REBICH HESPANHA, S. und HESPANHA, J.P. (2011): *Text visualization toolbox (TVT). A MATLAB toolbox to visualize large corpus of documents*.
- REGNAULD, N. und MCMASTER, R.B. (2007): A synoptic view of generalisation operators. In: MACKNESS, W.A., et al. (eds.): *Generalisation of geographic information: Cartographic modelling and applications*. Amsterdam, Boston.
- ROBINSON, A., H., MORRISON, J.L., MUEHRCKE, P.C., KIMERLING, J.A. und GUPTILL, S.C. (1978): *Elements of cartography*.
- RODGERS, P. (2005): Graph drawing technics for geographic visualization. In: DYKES, J., et al. (eds.): *Exploring Geovisualization*. Amsterdam: 143-158.
- ROZENBLAT, C. (2004): *Tissus de villes. Réseaux et systèmes urbains en Europe*. Habilitation à diriger des recherches, Université Paul Valéry - Montpellier III.
- ROZENBLAT, C. (2010): Opening the black box of agglomeration economies for measuring cities' competitiveness through international firm networks. *Urban Studies*. 47(13): 2841-2866.
- ROZENBLAT, C. und MELANÇON, G. (2009): A small world perspective on urban systems. In: BAVAUD, F. und MAGER, C. (eds.): *Handbook of theoretical and quantitative geography*. Lausanne: 431-467.
- ROZENBLAT, C., MÉLANÇON, G., BOURQUI, R. und AUBER, D. (2010): Comparing multilevel clustering methods on weighted graphs. the case of worldwide air passenger traffic 2000-2004. In: ROZENBLAT, C. und MÉLANÇON, G. (eds.): *Multilevel Analysis and visualization of Geographical Networks*. Berlin: in Druck.
- ROZENBLAT, C. und PUMAIN, D. (2007): Firm linkages, innovation and the evolution of urban systems. In: TAYLOR, P.J. (eds.): *Cities in Globalization: Practices, policies and theories*. London, New York: 130-156.
- SALES-PARDO, M., GUIMERÀ, R., MOREIRA, A.A. und AMARAL, L.A.N. (2007): Extracting the hierarchical organization of complex systems. *PNAS*. 104(39): 15224-15229.
- SALVINI, M.M., GNOS, A. und FABRIKANT, S.I. (2011): Cognitive plausible visualization of network data. *International cartographic conference (ICC)*, Paris.
- SASSEN, S. (1991): *The global city*. New York, London, Tokyo. Princeton.
- SASSEN, S. (1994): *Cities in a world economy*. Thousand Oaks.
- SASSEN, S. (1995): On concentration and centrality in the global city. In: KNOX, P.L. und TAYLOR, P.J. (eds.): *World Cities in a World System*. Cambridge, New York: 63-76.

- SASSEN, S. (1996): Cities and communities in the global economy. *American Behavioral Scientist*. 39(5): 629-639.
- SASSEN, S. (2002): Locating cities on global circuits. *Environment and Urbanization*. 14(1): 13-30.
- SCOTT, A.J. (eds.) (2001): *Global city-regions. Trends, theories, policy*. Oxford.
- SCOTT, A.J., AGNEW, J., SOJA, E. und STORPER, M. (2001): Global city-regions. In: SCOTT, A.J. (eds.): *Global City-Regions. Trends, Theory, Policy*. Oxford: 11-32.
- SHAWN, M., BROWN, M.W., KLAVANS, R. und BOYACK, K. (2011): OpenOrd. An open-source toolbox for large graph layout. *SPIE Conference on Visualization and Data Analysis (VDA)*, San Francisco.
- SHORT, J.R., KIM, Y., KUSS, M. und WELLS, H. (1996): The dirty little secret of world cities research. Data problems in comparative analysis. *International Journal of Urban and Regional Research*. 20(4): 697-717.
- SKUPIN, A. (2000): From metaphor to method. Cartographic perspectives on information visualization. *Symposium on Information Visualization*, Salt Lake City.
- SKUPIN, A. (2003): A novel projection using an artificial neural network. *International Cartographic Conference (ICC)*, Durban.
- SKUPIN, A. und AGRAWAL, P. (2008): Introduction. What is a self-organizing map? In: PRAGYA, A. und SKUPIN, A. (eds.): *Self-organising Maps: Application in geographic information science*. Chicester: 1-12.
- SKUPIN, A. und ESPERBÉ, A. (2011): An alternative map of the United States based on an n-dimensional model of geographic space. *Journal of Visual Languages & Computing*. 22(4): 290-304.
- SKUPIN, A. und FABRIKANT, S.I. (2003): Spatialization Methods. A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science*. 30(2): 95-115.
- SKUPIN, A. und FABRIKANT, S.I. (2007): Spatialization. In: WILSON, J.P. und FOTHERINGHAM, A.S. (eds.): *The handbook of geographical information science*. Reston: 61-79.
- SLOCUM, T., SLUTER, R.S., KESSLER, F.C. und YODERM, S.C. (2004): A qualitative evaluation of maptime. A programm for exploring spatiotemporal point data. *Cartographica*. 39(3): 43-68.
- SMITH, D.A. und TIMBERLAKE, M. (1995a): Cities in global matrices. toward mapping the world-system's city system. In: KNOX, P.L. und TAYLOR, P.J. (eds.): *World cities in a world-system*. New York: 79-97.
- SMITH, D.A. und TIMBERLAKE, M. (1995b): Conceptualising and mapping the structure of the world system's city system. *Urban Studies*. 32(2): 287-302.

- SMITH, D.A. und TIMBERLAKE, M. (2001): World city networks and hierarchies, 1977-1997. An empirical analysis of global air travel links. *American Behavioral Scientist*. 44(10): 1656-1678.
- SOJA, E. (2000): *Postmetropolis. Critical studies of cities and regions*. Malden.
- STEYVERS, M. und GRIFFITHS, T. (2007): Probabilistic topic models. In: LANDAUER, T.K., et al. (eds.): *Handbook of latent semantic analysis*. Mahwah: 427-448.
- STRUBE, M. und PONZETTO, S.P. (2006): WikiRelate! Computing semantic relatedness using Wikipedia. *National Conference on Artificial Intelligence*, Boston.
- TAYLOR, P.A. (2001): Specification of the world city network. *Geographical Analysis*. 33(2): 181-194.
- TAYLOR, P.A. (2012): The interlocking network model. In: DERUDDER, B., et al. (eds.): *International handbook of globalization and world cities*. Northampton: 51-63.
- TAYLOR, P.J. (1997): Hierarchical tendencies amongst world cities. A global research proposal. *Cities*. 14(6): 323-332.
- TAYLOR, P.J. (2000): World cities and territorial states under conditions of contemporary globalization. *Political Geography*. 19(-): 5-32.
- TAYLOR, P.J. (2004): *World city network. A global urban analysis*. London.
- TAYLOR, P.J. (2005a): Leading world cities. Empirical evaluations of urban nodes in multiple networks. *Urban Studies*. 42(9): 1593-1608.
- TAYLOR, P.J. (2005b): New political geographies. Global civil society and global governance through world city networks. *Political Geography*. 24(-): 703-730.
- TAYLOR, P.J., CATALANO, G. und WALKER, D.R.F. (2002a): Exploratory analysis of the world city network. *Urban Studies*. 39(13): 2377-2394.
- TAYLOR, P.J., CATALANO, G. und WALKER, D.R.F. (2002b): Measurement of the world city network. *Urban Studies*. 39(13): 2367-2376.
- TAYLOR, P.J. und WALKER, D.R.F. (2001): World cities. A first multivariate analysis of their service complexes. *Urban Studies*. 38(1): 23-47.
- TAYLOR, P.J., WALKER, D.R.F., CATALANO, G. und HOYLER, M. (2002c): Diversity and power in the world city network. *Cities*. 19(4): 231-241.
- THIERSTEIN, A. und FÖRSTER, A. (eds.) (2008): *The image and the region. Making mega-city regions visible!* Baden.
- THIERSTEIN, A., LÜTHI, S., BENTLAGE, M. und WIESE, A. (2011): Interlocking firm networks and emerging Mega-City regions. The relational geography of the knowledge economy in Germany. *Workshop TU, Munich*.
- TOBLER, W. (1970): A computer movie simulation of urban growth in the Detroit region. *Economic Geography*. 46(2): 234-240.

- TVERSKY, B. (2000): Some ways that maps and diagrams communicate. In: FREKSA, C., et al. (eds.): *Spatial Cognition II, Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications* Berlin, New York: 72-79.
- UNITED NATIONS STATISTICS DIVISION (2006): *Population of capital cities and cities of 100,000 and more inhabitants* Aus: <http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm>, Zugriff: 24.8.2006.
- VOSS, J. (2006): Collaborative thesaurus tagging the Wikipedia way. *ArXiv Computer Science*.
- WALL, R. und VAN DER KNAAP, B. (2005): Netscape - Europe and the Evolving World City Network. *ERSA conference papers*,
- WALL, R. und VAN DER KNAAP, B. (2009): Centrality and Structure within Contemporary Worldwide Corporate Networks. *GaWC Research Bulletin*. 295.
- WALL, R. und VAN DER KNAAP, B. (2011): Sectoral differentiation and network structure within contemporary worldwide corporate networks. *Economic Geography*. 87(3): 267-308.
- WARE, C. (2004): *Information visualization. Perception for design*. San Francisco.
- WARE, C., PURCHASE, H.C., COLPOYS, L. und MCGILL, M. (2002): Cognitive measurements of graph aesthetics. *Information Visualization*. 1(2): 103-110.
- WASSERMAN, S. und FAUST, K. (2008): *Social network analysis. Methods and applications*. Cambridge, New York.
- WICHMANN MATTHIESSEN, C., WINKEL SCHWARZ, A. und FIND, S. (2010): World cities of scientific knowledge. Systems, networks and potential dynamics. *Urban Studies*. 47(9): 1879-1898.
- WIKIPEDIA CONTRIBUTORS (2011a): *Category. Fundamental categories*. Aus: Zugriff: 15.05.2011.
- WIKIPEDIA CONTRIBUTORS (2011b): *Category. Main topic classifications*. Aus: http://en.wikipedia.org/wiki/Category:Fundamental_categories, Zugriff: 15.05.2011.
- WIKIPEDIA CONTRIBUTORS (2011c): *List of Wikipedias*. Aus: <http://s23.org/wikistats/wikipedias.html>, Zugriff: 27.06.2011.
- WIKIPEDIA CONTRIBUTORS (2011d): *Wikipedia*. Aus: <http://en.wikipedia.org/wiki/Wikipedia>, Zugriff: 21.06.2011.
- WIKIPEDIA CONTRIBUTORS (2011e): *Wikipedia. Relevanzkriterien*. Aus: <http://de.wikipedia.org/wiki/Wikipedia:RK>, Zugriff: 27.06.2011.
- WIKIPEDIA CONTRIBUTORS (2011f): *Wikipedians by location*. Aus: [http://en.wikipedia.org/wiki/Category:Wikipedians by location](http://en.wikipedia.org/wiki/Category:Wikipedians_by_location), Zugriff: 18.04.2011.

- WOLFE, J.M. und HOROWITZ, T.S. (2004): What attributes guide the deployment of visual attention and how do they do it? *Nature reviews. Neuroscience*. 5(-): 1-7.
- ZACHTE, E. (2011): *Wikistats. Wikimedia statistics*. Aus: <http://stats.wikimedia.org/FR/Sitemap.htm>, Zugriff: 18.04.2011.
- ZECH, T., GUREVYCH, I. und MÜHLHAUSER, M. (2007): Analyzing and accessing Wikipedia as a lexical semantic resource. *Biannual Conference of the Society for Computational Linguistics and Language Technology*, Tübingen.

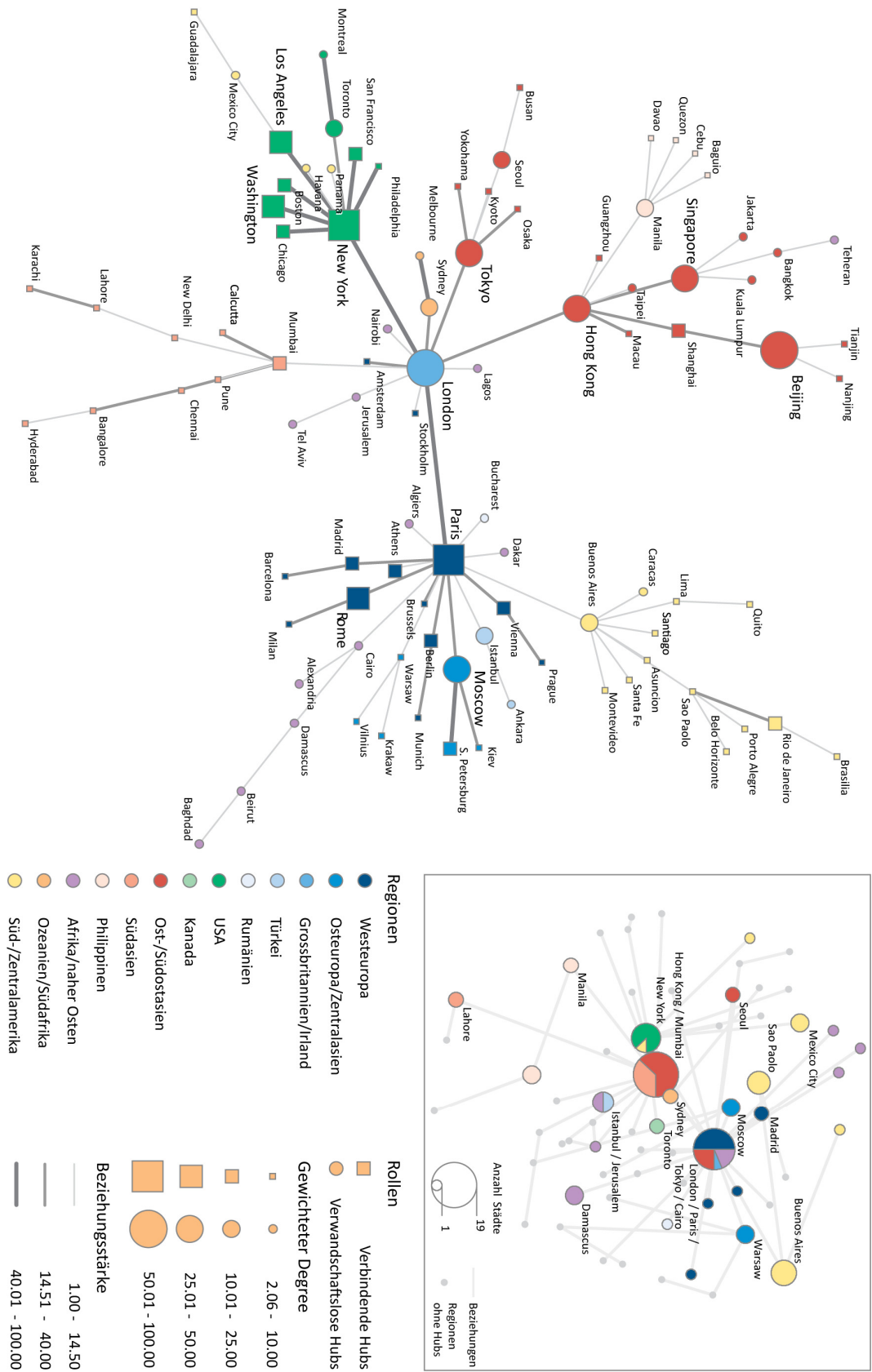
11 Anhang

Anhang 1:	Farbcodierung für Tabelle 4 bis Tabelle 7, Tabelle 27 und Anhang 9	215
Anhang 2:	Globales Stdtenetz zwischen den Hubs mit Hervorhebung der <i>betweenness</i> Zentralitt und deren Lage in der regionalen Ansicht	216
Anhang 3:	Funktionen im Stdtenetz und deren Grsse: Auszug GESELLSCHAFT/GESCHICHTE/ MILITR	217
Anhang 4:	Funktionen im Stdtenetz und deren Grsse: Auszug PERSONEN	218
Anhang 5:	Funktionen im Stdtenetz und deren Grsse: Auszug SPORT	219
Anhang 6:	Funktionen im Stdtenetz und deren Grsse: Auszug NATRLICHE/GEBAUTE UMWELT	220
Anhang 7:	Funktionen im Stdtenetz und deren Grsse: Auszug KUNST	221
Anhang 8:	Funktionen im Stdtenetz und deren Grsse: Auszug BILDUNG	222
Anhang 9:	Gegenberstellung der Resultate der Arbeit und des Forschungsstandes. Resultate: Rangierung der 95 Hubs im allgemeinen und in den funktionalen Netzwerken. Stand der Forschung: Kategorisierung und Rangierung der 95 Hubs nach den gleichen Quellen und Grundstzen wie in Kapitel 2.3 (siehe Tabelle 3 und Anhang 1).	223

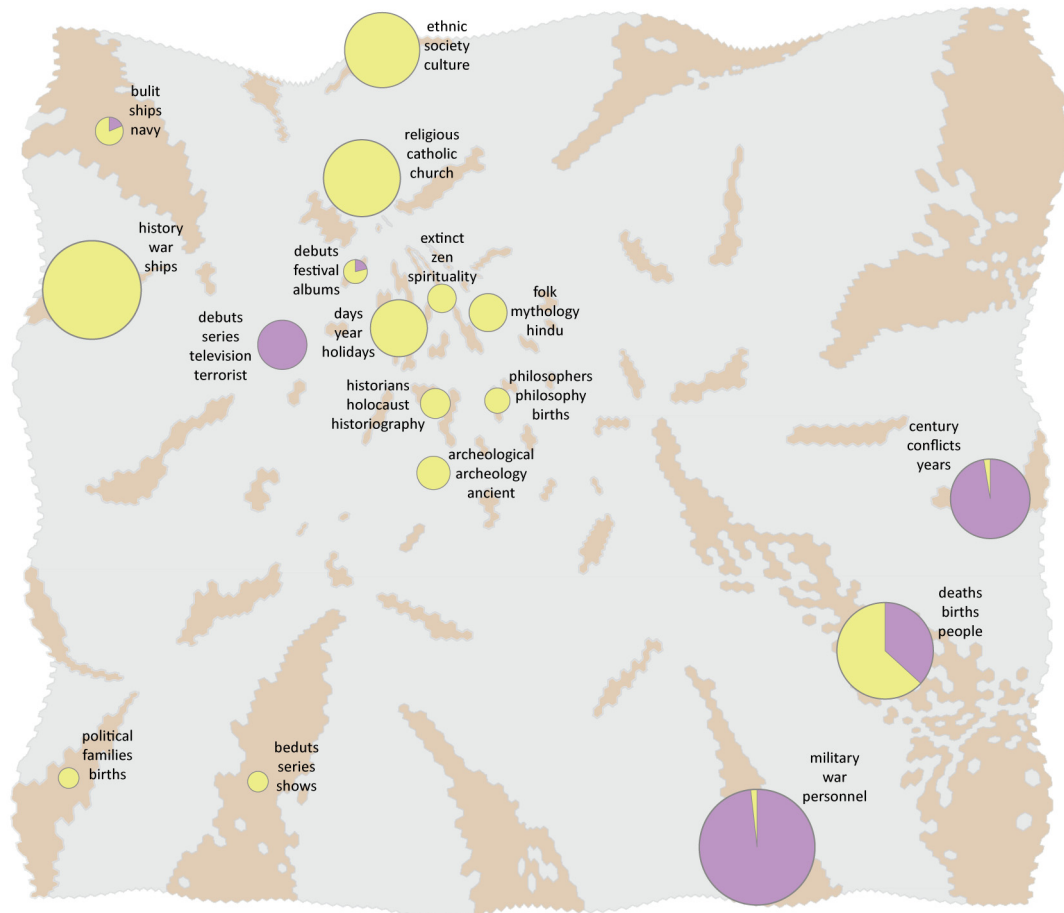
Anhang 1: Farbcodierung für Tabelle 4 bis Tabelle 7, Tabelle 27 und Anhang 9

Rang	Zentralitätsmasse			Motivation für die Klassengrenze
	generisch	OD	ID	
1-3				Anzahl Gfa nach Friedmann (1995)
4-15				Anzahl Städte in Alderson et al. (2010)
16-30				Anzahl Städte in Friedmann und Wall und van der Knaap (2009)
30-50				Anzahl Städte in Alderson und Beckfield (2004)
>50 / k.a.				

Anhang 2: Globales Städtenetz zwischen den Hubs mit Hervorhebung der *betweenness* Zentralität und deren Lage in der regionalen Ansicht



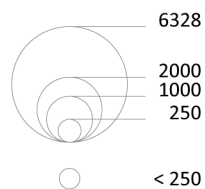
Anhang 3: Funktionen im Städtenetz und deren Grösse: Auszug GESELLSCHAFT/GESCHICHTE/
MILITÄR



Regionen mit Akteure

keine Daten

Anzahl Akteuren



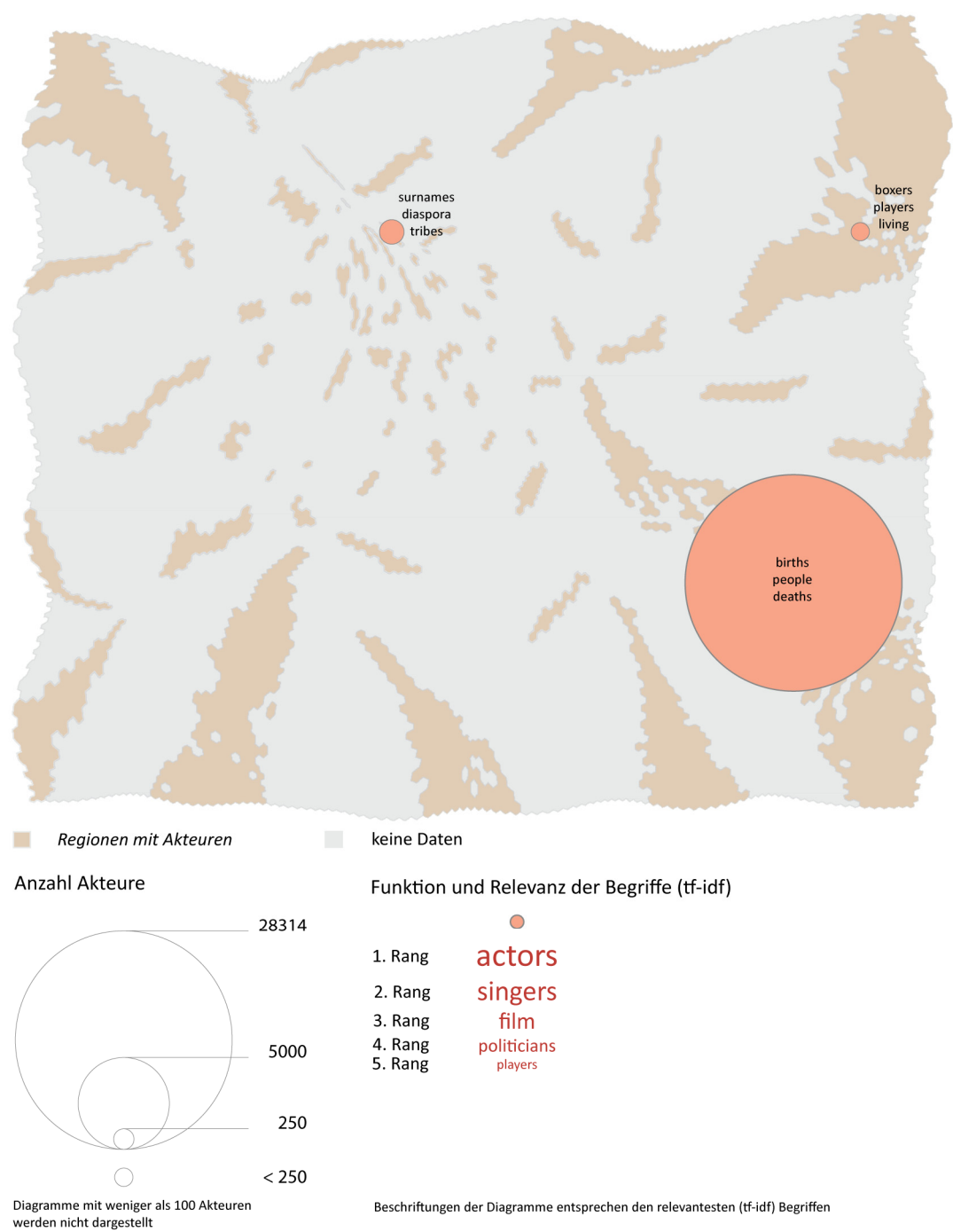
Sektoren mit weniger als 40 und Diagramme mit weniger als 100 Akteuren werden nicht dargestellt

Subfunktionen und Relevanz der Begriffe (tf-idf)

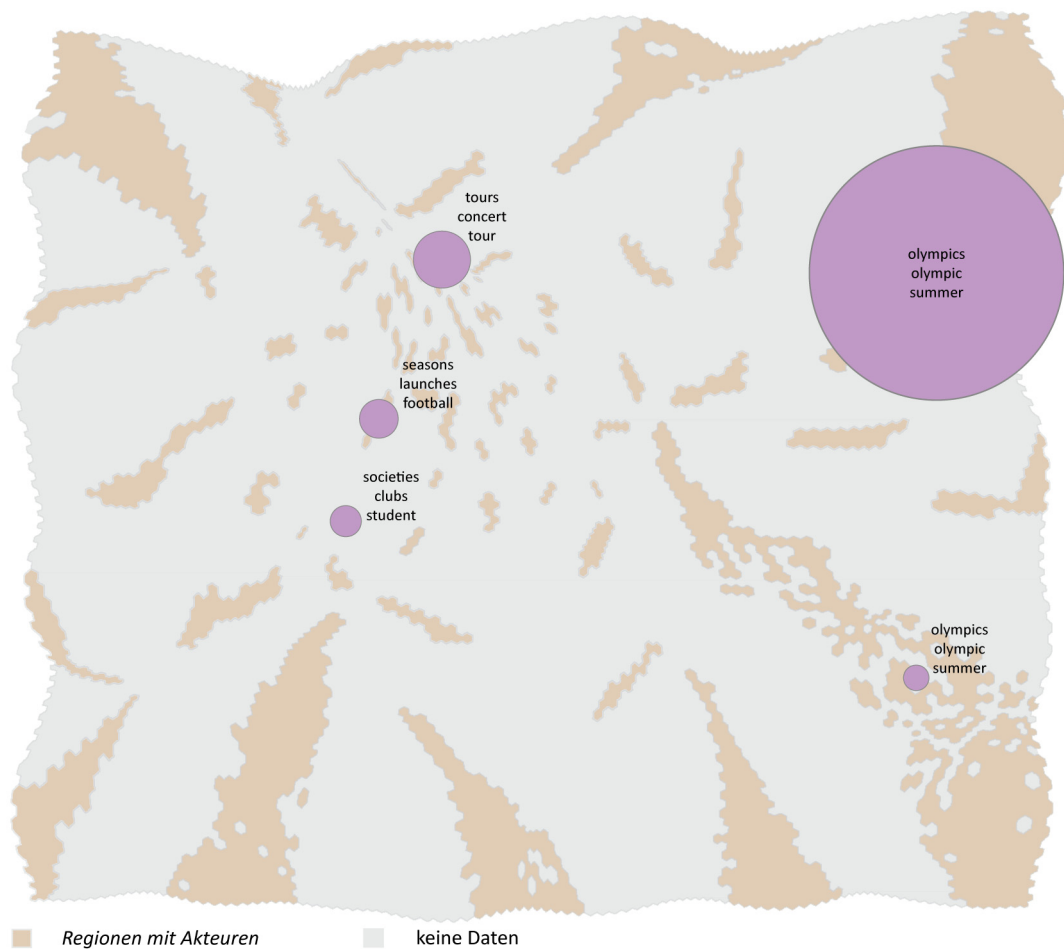


Beschriftungen der Diagramme entsprechen den relevantesten (tf-idf) Begriffen

Anhang 4: Funktionen im Städtenetz und deren Grösse: Auszug PERSONEN



Anhang 5: Funktionen im Städtenetz und deren Grösse: Auszug SPORT



Anzahl Akteure

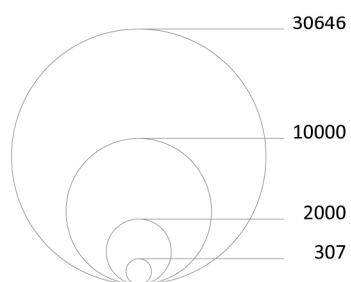


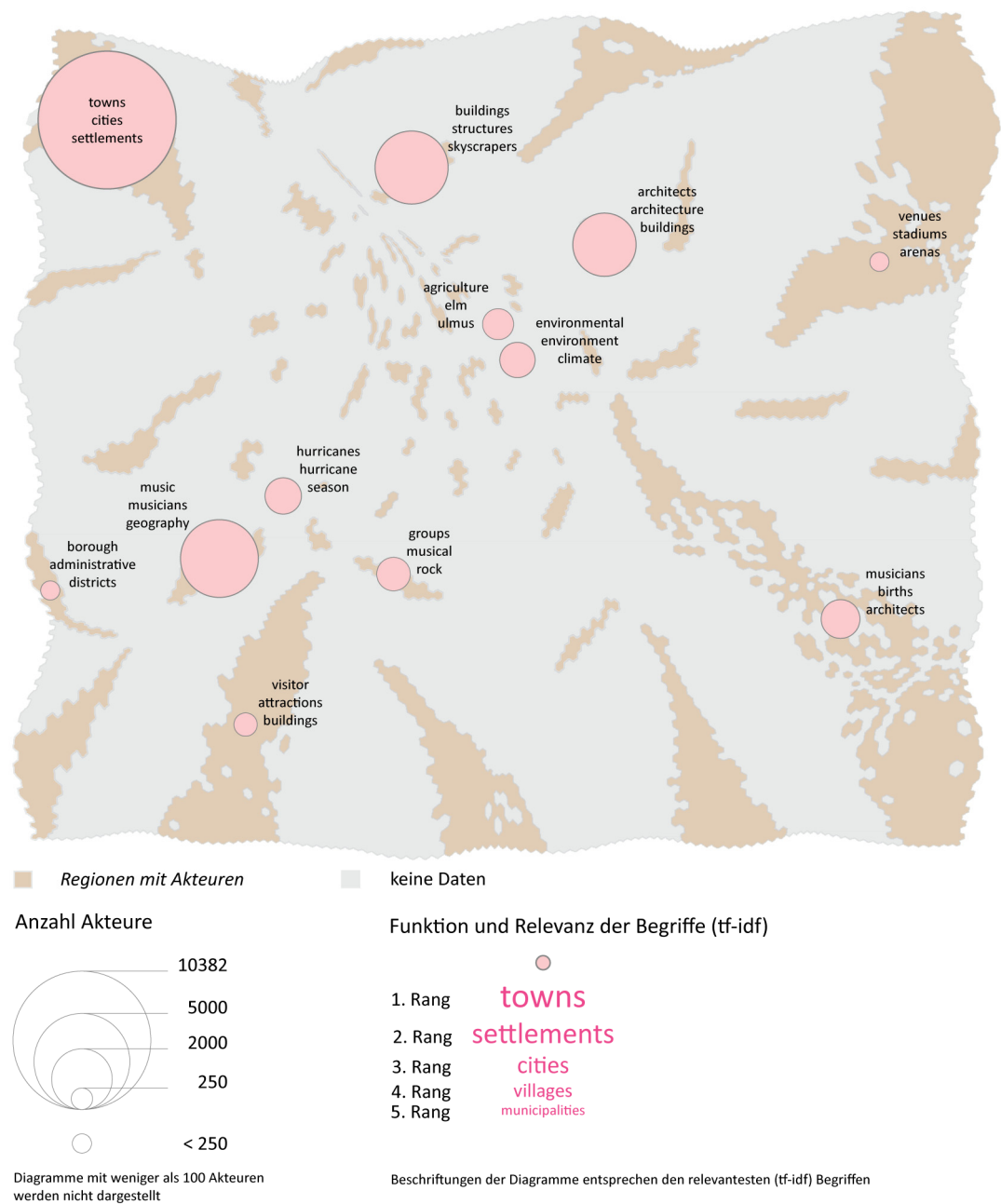
Diagramme mit weniger als 100 Akteuren werden nicht dargestellt

Funktion und Relevanz der Begriffe (tf-idf)

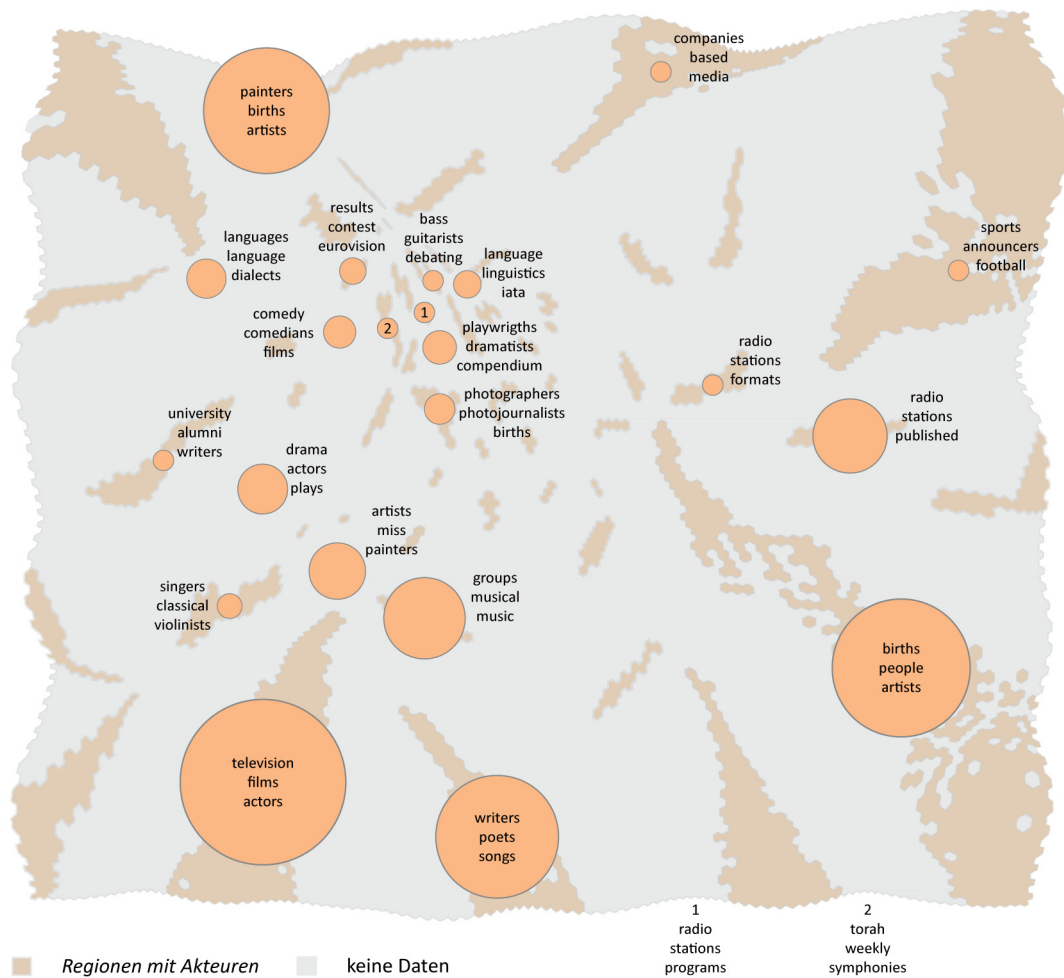


Beschriftungen der Diagramme entsprechen den relevantesten (tf-idf) Begriffen

Anhang 6: Funktionen im Städtenetz und deren Grösse: Auszug NATÜRLICHE/GEBAUTE UMWELT



Anhang 7: Funktionen im Städtenetz und deren Grösse: Auszug KUNST



Anzahl Akteure

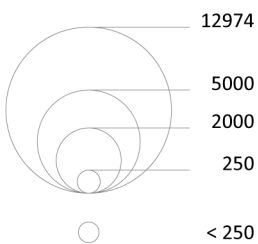


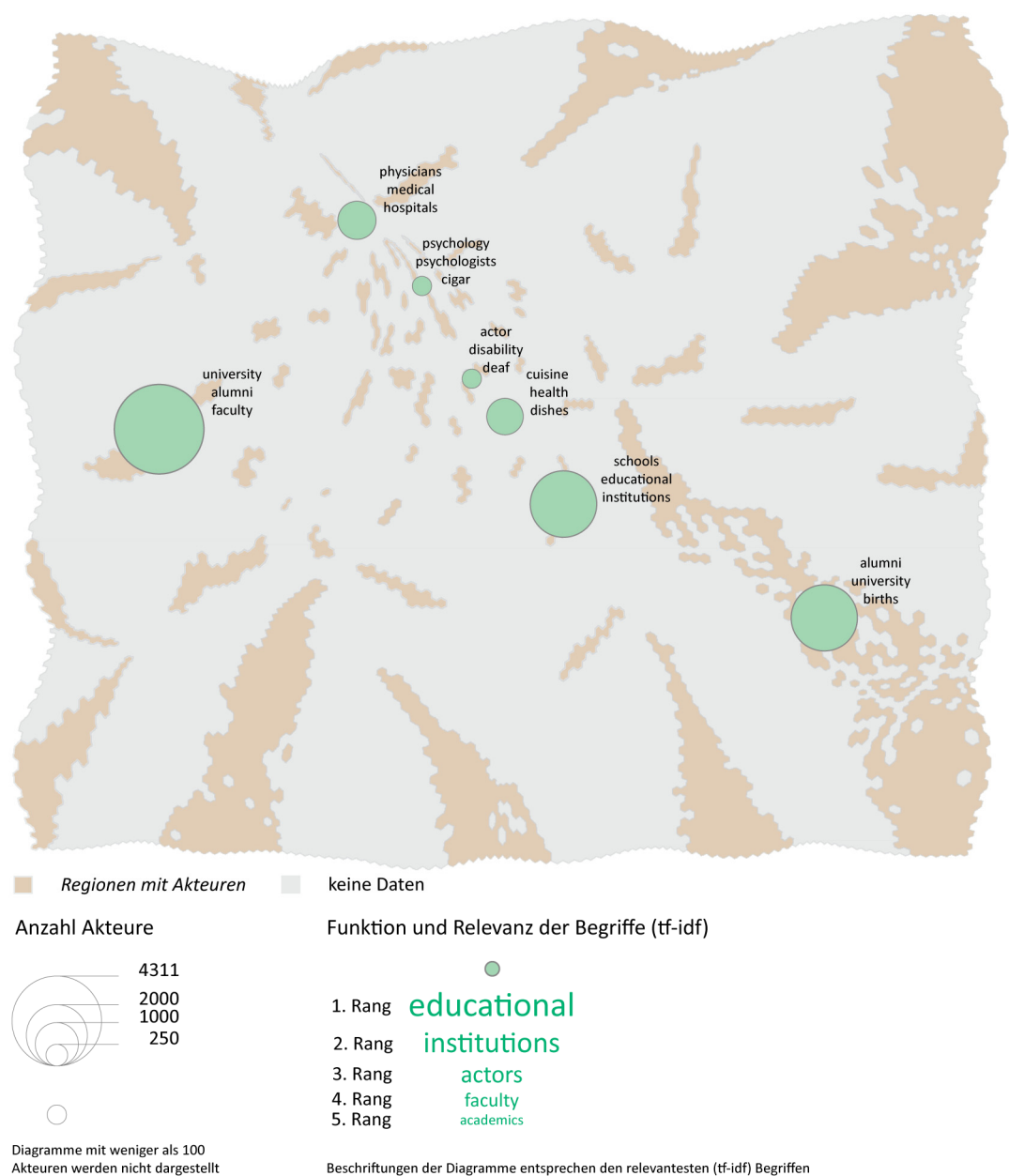
Diagramme mit weniger als 100 Akteuren werden nicht dargestellt

Funktion und Relevanz der Begriffe (tf-idf)

1. Rang actors
2. Rang painters
3. Rang films
4. Rang film
5. Rang singers

Beschriftungen der Diagramme entsprechen den relevantesten (tf-idf) Begriffen

Anhang 8: Funktionen im Städtenetz und deren Grösse: Auszug BILDUNG



Anhang 9: Gegenüberstellung der Resultate der Arbeit und des Forschungsstandes. Resultate: Rangierung der 95 Hubs im allgemeinen und in den funktionalen Netzwerken. Stand der Forschung: Kategorisierung und Rangierung der 95 Hubs nach den gleichen Quellen und Grundsätzen wie in Kapitel 2.3 (siehe Tabelle 3 und Anhang 1).

Stadt	Eigene Resultate										Global Cities		World Cities			
	Allgemeines Netzwerk		Funktionale Netzwerke								GaWC-Forschungsgruppe	Wall und v.d. Knaap		Alderson und Beckfield		Friedmann
	3'532 Städte	95Hubs	Wirtschaft/Technologie	Gesellschaft/Geschichte/Militär	Personen	Gebaute/natürliche Umwelt	Kunst	Politik	Sport	Bildung		OD (2005)	ID (2005)	OD (2005)	ID (2005)	
											Global Connectivity (2010)					1995
London	1	1	2	1	1	2	2	1	6	2	α++	3	2	4	2	Gfa
New York	2	2	1	4	3	1	1	2	19	1	α++	1	1	1	1	Gfa
Paris	3	3	3	2	2	3	3	4	4	3	α+	2	5	2	5	lra
Los Angeles	4	4	9	12	4	11	4	6	9	5	α					Mna
Chicago	5	5	5	11	8	5	6	5	20	6	α+	14	19	20	29	Ra
Washington	6	6	8	5	7	4	10	3	44	4	α				23	
Berlin	7	7	18	9	6	6	5	7	13	7	β+	27				
Rome	8	8	24	3	10	12	9	8	11	14	β+				30	
San Francisco	9	11	7	8	11	17	7	11	48	10	α				10	Ra
Moscow	10	9	15	7	9	8	12	9	8	20	α					
Boston	11	14	6	13	14	31	11	10	34	9	α-		28		44	Ra
Vienna	12	12	26	6	5	13	8	12	29	8	α-			17	35	
Toronto	13	18	17	33	15	23	14	18	22	13	α		9	11	18	Ra
Philadelphia	14	16	13	16	13	22	13	14	30	11	α-	22			21	
Tokyo	15	10	11	24	17	10	15	20	1	17	α+	8	6	3	6	Gfa
Sydney	16	17	16	23	19	27	22	25	7	21	α+		23		25	INA
Beijing	17	15	14	19	34	7	32	15	3	18	α				24	
Montreal	18	22	28	35	21	20	24	17	15	19	β+			25	50	Ra
Munich	19	20	25	20	16	54	16	35	14	15	α-	13	26	7	28	Ra
Hong Kong	21	13	4	17	18	15	19	13	36	12	α+		4		4	Ra
Madrid	22	21	23	22	23	14	20	19	24	30	α		13		8	lra
Athens	23	23	45	14	45	28	37	30	2	45	β+					
Melbourne	26	24	22	38	26	46	25	37	12	32	α-			21	38	
Milan	27	26	19	21	30	50	18	39	25	43	α		7		9	Ra
Stockholm	28	30	39	44	28	16	33	26	16	48	β+				14	
Barcelona	29	27	29	40	37	52	26	27	10	37	α-			21	40	Mna
Brussels	30	25	27	26	27	9	23	16	45	31	α	9	12	17	7	9
Amsterdam	31	32	20	31	29	64	21	33	32	24	α	7	8	9	16	Mna
Prague	32	34	47	30	22	34	28	34	31	29	β+					
Singapore	33	19	10	27	24	21	27	22	27	16	α+	30	3		3	11

	Eigene Resultate										Global Cities		World Cities								
	Allgemeines Netzwerk		Funktionale Netzwerke								GaWC-Forschungsgruppe	Wall und v.d. Knaap		Alderson und Beckfield				Friedmann			
Stadt	3'532 Städte	95Hubs	Wirtschaft/Technologie	Gesellschaft/Geschichte/Militär	Personen	Gebaute/natürliche Umwelt	Kunst	Politik	Sport	Bildung		Global Connectivity (2010)	OD (2005)	ID (2005)	OD (2005)	ID (2005)	OD (2000)		ID(2000)	OD (2007)	ID (2007)
St.Petersburg	34	28	32	15	12	49	17	24	42	40	γ										
Seoul	43	29	35	48	64	24	44	45	5	39	α		23		30	31	29				Ina
Mexico City	44	36	44	46	36	30	31	44	23	34	α		12		16		16				INA
Buenos Aires	45	31	40	43	31	19	29	28	18	47	α				22		28				
Warsaw	47	37	41	32	20	26	30	38	58	42	α-										
Mumbai	51	35	21	25	33	32	39	21	55	23	α										
Shanghai	53	33	12	29	38	29	36	23	51	25	α+										
Istanbul	56	39	37	18	32	38	38	36	47	46	α-										
RiodeJaneiro	64	38	34	37	41	51	35	41	26	35	β-						46				
Bangkok	68	40	31	62	50	40	53	49	17	52	α-		11		18		25				
São Paulo	70	41	30	51	46	39	34	57	35	44	α		22		20		18				INA
Cairo	74	42	52	28	35	18	41	40	53	49	β+										
Chennai	82	46	36	45	55	56	54	29	59	22	β										
Manila	85	45	33	42	48	37	45	56	52	50	β+						42				
Calcutta	88	48	42	39	43	65	49	32	57	28	β-										
New Delhi	89	44	51	69	56	25	48	31	50	36	α-										
Bucharest	90	53	58	65	49	42	52	42	61	59	β										
Osaka	91	47	48	68	59	73	43	78	21	57	β-	21		28		9	39	13			Ra
KualaLumpur	92	43	46	76	60	33	56	53	28	60	α						32				
Kiev	98	51	54	52	42	45	42	47	65	69	β										
Jerusalem	107	49	88	10	25	71	40	46	90	41											
Havana	111	54	68	61	52	36	50	60	56	73											
Santiago	113	50	59	59	57	48	62	68	33	64	α-										
Tehran	118	58	72	75	53	47	64	61	46	71											
Jakarta	120	52	49	73	70	43	65	66	43	67	α		15		15		31				
Bangalore	125	57	38	56	74	68	63	43	66	26	β+										
Tel Aviv	135	56	60	55	39	44	47	50	72	65	β+										
Lima	139	55	67	66	71	60	67	58	37	62	β										
Kraków	140	66	75	36	40	77	46	70	88	55											
Caracas	147	60	61	84	65	53	60	71	49	72	β						45				
Karachi	149	67	56	57	61	69	74	73	73	27	β										
Beirut	150	59	62	49	44	59	57	52	62	56											
Guangzhou	153	62	50	60	68	58	70	55	64	53	β										
Lahore	156	81	73	71	63	86	75	77	76	33	γ+										

	Eigene Resultate										Global Cities		World Cities										
	Allgemeines Netzwerk		Funktionale Netzwerke								GaWC-Forschungsgruppe	Wall und v.d. Knaap		Alderson und Beckfield		Friedmann							
Stadt	3'532 Städte	95Hubs	Wirtschaft/Technologie		Gesellschaft/Geschichte/Militär		Personen	Gebaute/natürliche Umwelt		Kunst	Politik	Sport	Bildung	Global Connectivity (2010)	OD (2005)	ID (2005)	OD (2005)	ID (2005)	OD (2000)	ID(2000)	OD (2007)	ID (2007)	1995
Ankara	160	70	78	82	72	41	78	48	84	68													
Taipei	162	63	43	80	75	61	55	62	68	51	α-						26		37				
Baghdad	165	72	89	53	54	66	61	59	83	80													
Damascus	166	65	81	41	58	55	76	64	69	81													
Vilnius	168	73	86	50	51	72	58	72	78	76	γ												
Santa Fe	171	68	66	72	73	63	72	74	60	63													
Yokohama	172	64	65	64	47	90	73	84	40	84													
Busan	173	71	69	81	86	80	94	83	41	77													
Panama	178	78	53	54	67	92	77	69	79	74	β-												
Hyderabad	184	75	55	83	77	76	79	51	74	38	γ+												
Montevideo	190	61	71	67	66	70	69	65	38	87	β												
Algiers	199	79	77	70	81	57	82	80	75	89													
Nairobi	203	83	76	86	79	62	88	67	82	66	γ+												
Alexandria	205	77	74	34	62	87	66	75	85	75													
Brasília	209	74	80	89	87	35	80	76	77	82													
Pune	216	82	64	77	76	75	71	63	91	54	γ-												
Macau	219	69	57	58	69	78	59	54	67	61													
Kyoto	240	80	84	47	85	89	51	89	80	58													
Dakar	246	87	83	78	89	67	92	85	87	88													
Guadalajara	255	86	90	85	78	91	68	87	63	83	γ												
Tianjin	256	85	63	63	80	82	93	81	71	78													
Nanjing	272	88	70	74	82	83	85	79	92	79													
Quito	274	84	79	87	83	74	89	86	54	91	γ												
Asunción	281	76	92	88	90	79	91	82	39	92													
Belo	297	90	82	92	92	85	86	95	81	86													
Lagos	299	92	87	79	91	81	95	90	89	95	β-												
Porto Alegre	317	89	85	90	88	84	83	91	70	85	γ-												
Quezon	331	91	91	91	84	88	84	88	86	70													
Davao	444	93	94	93	93	93	81	94	94	90													
Cebu	490	94	93	94	94	94	87	92	93	93													
Baguio	567	95	95	95	95	95	90	93	95	94													

Curriculum Vitae

SALVINI Marco Michele

geboren am 9. November 1983, von Cama / GR

Ausbildung

Universität Zürich, Zürich

Sept. 2008 – Dez. 2012 Promotionsstudium MNF

Dissertation *Spatialization* von nutzergenerierten Inhalte für die explorative Analyse des globalen Städtennetzes. Betreuung und Leitung: Prof. Dr. Sara Irina Fabrikant. Promotionskomitee: Prof. Dr. C. Rozenblat, Universität Lausanne; Dr. T. Reichenbacher, Universität Zürich

Sept. 2006 – Mai 2008 Master of Science in Geographie

Masterarbeit Kreative räumliche Allokation. Wohnstandortwahl der *Creative Class* in der Agglomeration Zürich. Betreuung: Dr. Corinna Heye. Leitung: Prof. Dr. Hans Elsasser

Nebenfach Raumplanung und Landschaftsentwicklung, Institut für Raum- und Landschaftsentwicklung, ETHZ

Okt. 2003 – Mai 2006 Bachelor of Science in Geographie

Bachelorarbeit Sozialer Wohnungsbau in der Schweiz und in den Niederlanden. Betreuung: Dr. Karin Schwiter. Leitung: Prof. Dr. Hans Elsasser.

Nebenfächer Volkswirtschaftslehre und Geschichte der Neuzeit

Liceo Cantonale Bellinzona, Bellinzona

Sept. 1999 – Jun. 2003 Eidgenössische Matura mit Schwerpunktfach Biologie und Chemie

Maturarbeit Identità nazionale Elvetica. Incontro e scontro con la globalizzazione.
Betreuung: A. Codoni, N. Willis

Publikationen

Bruggmann, A., Salvini, M.M., Fabrikant, S.I. (2013). Cartograms of self-organizing maps to explore user generated content. *Proceedings, 26th International Cartographic Conference*, Dresden, Deutschland (abstract eingereicht)

Salvini, M.M., Fabrikant, S.I. (2012). Scale-dependent validation framework for spatialization. *Proceedings, 7th International Conference on Geographic Information Science*, Columbus, Ohio, USA

Pospischil, A., Hässig, M., Vogel, R., Salvini, M., Fabrikant, S.I., Axhausen, K., Schenker, N.S., Erni, D., Guscetti, F. (2012): Hundepopulation und Hunderassen in der Schweiz von 1955 bis 2008, *Schweizer Archiv für Tierheilkunde* (angenommen)

Salvini, M.M., Gnos, A.U., FABRIKANT, S.I. (2011). Cognitively Plausible Visualization of Network Data. *Proceedings, 25th International Cartographic Conference*, Paris, Frankreich

Fabrikant, S.I., Salvini, M.M. (2011). Charting the ICA World of Cartography 1999–2009. *Proceedings, 25th International Cartographic Conference*, Paris, Frankreich

Salvini, M. M., Heye C. (2008): Wohnstandortwahl der Creative Class in der Agglomeration Zürich. *disP*, 175(8), 26-39